DOCUMENT RESUME

ED 183 569 TM 010 024

AUTHOR	Wilson, Kenneth M.
TITLE	The Validation of GRE Scores as Predictors of First-Year Performance in Graduate Study: Report of the GRE Cooperative Validity Studies Project. GRE Board Research Report GREB No. 75-8R.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	GREB-75-8R
PUB DATE	Jun 79
NOTE	163p.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Academic Achievement: *College Entrance Examinations: Departments: Females: Foreign Students: *Grade Point Average: *Graduate Study: Higher Education: Interagency Cooperation: Minority Groups: Models: *Predictive Ability (Testing): *Predictive Validity: Predictor Variables: Research Projects: Testing Programs: Undergraduate Study: Weighted Scores
IDENTIFIERS	*Graduate Record Examinations: GRE Cooperative Validity Studies Project

ABSTRACT

The Graduate Record Examinations (GRE) Cooperative Validity Studies Project began in 1975 to fill the need for cooperation between graduate schools and testing agencies and for graduate level validity studies. More than 150 data sets from the 39 participating schools, representing over 19 fields of study, were analyzed. The data confirmed earlier studies which supported the correlational validity for GRE Aptitude and Advanced tests and undergraduate grade point average (GPA) with the grade point averages of first-time, full-time graduate students. Findings indicated recurring problems which must be solved in graduate-level validity studies: small samples, unstable weighting, range restriction, and criterion selection. However, it was determined that basic validity studies can be conducted, despite these problems. Limited evidence was also provided for the validity of GRE tests for females, minorities, and foreign students. It was concluded that recurring participation in these kinds of studies is more important in graduate programs with small departmental samples as units of analysis than in programs with large entering classes. (Selected findings, survey instruments, and a list of participating schools are appended.) (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183569

TM010 024

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

GRE

GRADUATE RECORD EXAMINATIONS

THE VALIDATION OF GRE SCORES AS PREDICTORS OF FIRST-YEAR PERFORMANCE IN GRADUATE STUDY: REPORT OF THE GRE COOPERATIVE VALIDITY STUDIES PROJECT

Kenneth M. Wilson

GRE Board Research Report GREB No. 75-8R

June 1979

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
ETS

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

GRE BOARD RESEARCH REPORTS
FOR GENERAL AUDIENCE

Altman, R. A. and Wallmark, M. M. A Summary of Data from the *Graduate Programs and Admissions Manual*. GRE Board Research Report GREB No. 74-1R, January 1975.

Baird, L. L. An Inventory of Documented Accomplishments. GRE Board Research Report GREB No. 77-3R, June 1979.

Baird, L. L. Cooperative Student Survey (The Graduates [$2.50 each], and Careers and Curricula). GRE Board Research Report GREB No. 70-4R, March 1973.

Burns, R. L. Graduate Admissions and Fellowship Selection Policies and Procedures (Part I and II). GRE Board Research Report GREB No. 69-5R, July 1970.

Centra, J. A. How Universities Evaluate Faculty Performance: A Survey of Department Heads. GRE Board Research Report GREB No. 75-5bR, July 1977. ($1.50 each)

Centra, J. A. Women, Men and the Doctorate. GRE Board Research Report GREB No. 71-10R, September, 1974. ($3.50 each)

Clark, M. J. The Assessment of Quality in Ph.D. Programs: A Preliminary Report on Judgments by Graduate Deans. GRE Board Research Report GREB No. 72-7aR, October 1974.

Clark, M. J. Conditions Influencing the Career Accomplishments of Recent Ph.D.'s. GRE Board Research Report GREB No. 76-2R, June 1979.

Clark, M. J. Program Review Practices of University Departments. GRE Board Research Report GREB No. 75-5aR, July 1977. ($1.00 each)

Donlon, T. F. Annotated Bibliography of Test Speededness. June 1979.

Flaugher, R. L. The New Definitions of Test Fairness In Selection: Developments and Implications. GRE Board Research Report GREB No. 72-4R, May 1974.

Fortna, R. O. Annotated Bibliography of the Graduate Record Examinations. July 1979.

Frederiksen, N. and Ward, W. C. Measures for the Study of Creativity in Scientific Problem-Solving. May 1978.

Hartnett, R. T. and Willingham, W. W. The Criterion Problem: What Measure of Success in Graduate Education? GRE Board Research Report GREB No. 77-4R, March 1979.

Knapp, J. and Hamilton, I. B. The Effect of Nonstandard Undergraduate Assessment and Reporting Practices on the Graduate School Admissions Process. GRE Board Research Report GREB No. 76-14R, July 1978.

Lannholm, G. V. and Parry, M. E. Programs for Disadvantaged Students in Graduate Schools. GRE Board Research Report GREB No. 69-1R, January 1970.

Reilly, R. R. Critical Incidents of Graduate Student Performance. GRE Board Research Report GREB No. 70-5R, June 1974.

Rock, D. A. The Prediction of Doctorate Attainment in Psychology, Mathematics and Chemistry. GRE Board Research Report GREB No. 69-6aR, June 1974.

Schrader, W. B. Admissions Test Scores as Predictors of Career Achievement in Psychology. GRE Board Research Report GREB No. 76-1R, September 1978.

Wild, C. L. Summary of Research on Restructuring the Graduate Record Examinations Aptitude Test. February 1979.

Wilson, K. M. The GRE Cooperative Validity Studies Project. GRE Board Research Report GREB No. 75-8R, June 1979.

Wiltsey, R. G. Doctoral Use of Foreign Languages: A Survey. GRE Board Research Report GREB No. 70-14R, 1972. (Highlights $1.00, Part I $2.00, Part II $1.50).

Witkin, H. A.; Moore, C. A.; Oltman, P. K.; Goodenough, D. R.; Friedman, F.; and Owen, D. R. A Longitudinal Study of the Role of Cognitive Styles in Academic Evolution During the College Years. GRE Board Research Report GREB No. 76-10R, February 1977 ($5.00 each).

The Validation of GRE Scores as Predictors of First-Year Performance in Graduate Study: Report of the GRE Cooperative Validity Studies Project

Kenneth M. Wilson

GRE Board Research Report GREB No. 75-8R

June 1979

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

Copyright © 1979 by Educational Testing Service. All rights reserved.

Acknowledgments

The GRE Cooperative Validity Studies Project has been a collaborative undertaking, involving the cooperative interaction of members of the faculty and administration of a large number of graduate schools, members of the ETS staff, and the Graduate Record Examinations Board. The project was made possible by the sustained support and encouragement of the Graduate Record Examinations Board. That cooperation, support, and encouragement is acknowledged with sincere appreciation.

Members of the GRE Program staff, especially Messrs. Robert Altman, Bernard Khoury, and Ray Thompson, offered advice, helpful suggestions, and encouragement. Ms. Mary Jo Clark provided helpful advice on a variety of research and administrative questions during the entire project. Programming assistance provided by Messrs. James Ferris and Richard Harrison was essential to the project. Assistance in processing data for the project was provided by Ms. Faith Thompson.

The contribution of those named specifically and others who contributed to the project is gratefully acknowledged.

Kenneth M. Wilson

CONTENTS

		Page
Part I	Context and Perspective	3
	A Cooperative Studies Rationale	5
Part II	Developing and Implementing Cooperative Studies	6
Part III	General Characteristics and Principal Findings of the Cooperative Studies.	11
	Characteristics of Samples and Data	11
	Principal Study Findings.	14
	Patterns of Validity Coefficients by Field. . . .	18
	Comparison with Other Validity Study Findings . .	22
	Validity for Subgroups.	27
Part IV	The Problem of Combining Predictors in Small Graduate Departments.	37
	Testing a "Common Weights" Hypothesis	39
	Regression Results When Data were Pooled.	42
	Implications of the Findings.	47
Part V	Retrospect and Prospect	49
References	. . . .	53
Appendix A:	GRE Program Related Efforts to Promote Validation Research: Review and Appraisal	57
Appendix B-1:	Forms Used in Survey of Deans of CGS Member Graduate Schools.	73
B-2:	Survey of Graduate School Validity Study Activities and Interests: Summary of Findings. .	85
B-3:	Basic One-year Validity Study Model Used in the Cooperative Studies	95
B-4:	Brief Description of Selected Studies Using a Two-year Study Model.	105

CONTENTS

Page

Appendix C-1: Illustrative Institutional Report: Graduate School A. 117

C-2: Tabular Summary of Selected Validity Study Findings, March 1978. 139

Appendix D: Examination of Departmental Samples with Deviant Weights for Predictors. 157

THE GRE COOPERATIVE VALIDITY STUDIES PROJECT

Kenneth M. Wilson
Educational Testing Service

Decisions to admit some applicants for graduate study and to reject others have serious implications for individuals, graduate schools, and society. It is of the utmost importance that all such decisions should be guided by up-to-date and reliable knowledge regarding the predictive validity of data employed in screening applicants for admission. All parties to the development of the Graduate Record Examinations (GRE) Program, from the outset, have recognized the need for empirical evidence regarding the predictive validity of GRE tests and other preadmissions variables.

Concern for predictive validity has been expressed directly in the form of GRE-Board support for a variety of ad hoc projects that have had the general aim of helping to increase the amount and quality of information available about the validity of GRE Aptitude and Advanced Tests, and to improve the validity study process generally.

Projects undertaken by ETS, at the suggestion of either the Committee on Testing of the Association of Graduate Schools (AGS) or the Graduate Record Examinations Board, have been of three principal types:*

1. those concerned with collecting and disseminating information and insights gained from locally conducted institutional/departmental validity studies (e.g., Lannholm and Schrader, 1951; Lannholm, 1960, 1968, 1972);
2. those concerned with conducting centrally planned validity studies with the cooperation of selected graduate schools and/or departments (e.g., Lannholm, Marco, & Schrader, 1968; Boldt, 1975); and
3. those concerned with the study of particular applied, methodological, or conceptual aspects of the validation process or with particular problems and issues (e.g., Boldt, 1975; Carlson, Evans, & Kuykendall, 1973; Reilly, 1971, 1974; Rock & Harmon, 1972).

These diverse approaches have sought to encourage and improve validation research by improving the scope and quality of information available to graduate schools regarding the validity of GRE tests and the different ways in which validity studies might be carried out; by actually conducting studies using data provided by selected institutions or departments; and by focusing attention on important validation research problems and exploring promising developments for coping with these problems.

---

* A detailed review of representative projects is provided in Appendix A.

Despite the impetus represented by these projects, spanning a full quarter of a century, the number of graduate level validity studies has remained low. The Cooperative Validity Studies Project described in this report reflects an extension and intensification of efforts on the part of the Graduate Record Examinations Board to promote and facilitate the participation of graduate schools and departments in the GRE validity-study process.

An immediate objective of the project was to enlist the cooperation of graduate schools and departments in studies designed to obtain up-to-date information regarding the predictive validity of GRE Aptitude and Advanced Tests and other variables used in admission (such as the Undergraduate Grade Point Average or UPGA) with respect to relevant criteria of performance in graduate study. It was assumed that experience gained in developing and implementing cooperative studies with a wide range of graduate schools and departments would contribute to the development of longer term arrangements through which the GRE Program might facilitate the recurring participation of schools and departments in GRE validation research.

This report provides an overview of the Cooperative Validity Studies Project. The principal project activities, findings, and conclusions are described following a brief examination of some of the reasons for (a) the comparatively low level of validity study activity at the graduate level and (b) the need for cooperative interaction between graduate schools and departments, the GRE Board, and the GRE Program at ETS in validation research.

PART I: CONTEXT AND PERSPECTIVE

The volume, scope, and coverage of validity study activity in graduate school settings has been quite low both in relation to the number of settings in which validity studies could and should be conducted (e.g., all graduate departments in which GRE scores may affect admissions decisions) and as compared with the volume of validity study activity in other educational settings where circumstances have been more conducive to the widespread application of standard validity study models and procedures.

In undergraduate and law school settings, for example, the validity study process has been faciliated by (a) the existence of comparatively large entering cohorts of students engaged in comparable academic pursuits, especially during the initial phases of their educational programs, and (b) the general acceptance of one performance index--namely, the first-year grade point average--as the criterion against which to validate admissions variables. Questions regarding the predictive validity of individual admissions variables and the most effective combinations of those variables for predicting first-year grades are addressed systematically by applying standard statistical models, principally multiple regression analysis, to data for sizeable samples typically corresponding to an entering first-year class for each institution (Schrader, 1971, 1977).

These conditions have been conducive to the regularization of institutional participation in validity studies employing standard design and methodology. Summaries of findings, prepared from time to time, provide normative perspective with respect to trends across institutions and over time in patterns of correlational validity for relevant predictors.

- In a recent review of law school validity studies, for example, Schrader (1977) drew upon the results of over 625 studies involving the Law School Admission Test and an undergraduate grade point average, in relation to a first-year grade average criterion, completed for 150 law schools between 1948 and 1975.

- The number of validity studies in undergraduate settings is also high. During the period 1964-1968, for example, almost 1,900 validity studies involving College Board tests in relation to freshman-level grade point average criteria were completed at ETS alone (internal communication).

In graduate settings the situation is much more complex, and validation research has not become part of an established routine. Each graduate school has several "entering classes" each year, corresponding to distinctive subgroups definable, for example, in terms of field of study (department in which enrolled), type of degree program (e.g., terminal master's, master's-doctorate sequence, doctoral only), and educational status (e.g., first-time graduate student, master's-degree holder, etc.). These subgroups represent cohorts for which separate validity studies are likely to be needed, and each such cohort typically is quite small.

Problems related to small samples and therefore unstable estimates of parameters (underlying relationships between predictor and criterion variables), endemic in graduate school validation research, have been compounded by lack of a working consensus regarding the most useful (appropriate or relevant) criterion of "success" in graduate study (Willingham, 1974). The graduate grade point average reflects one relevant dimension of performance--and it has been the most consistently employed criterion in ad hoc validity studies--but it has failed to command the widespread acceptance accorded the first-year grade point average as a criterion in undergraduate and professional school settings.

The foregoing litany of deterrents helps to explain the fact that comparatively few graduate schools assess the validity of GRE tests, or other admissions variables, systematically and regularly. According to a GRE-sponsored survey of member institutions of the Council of Graduate Schools (Burns, 1970), 57 percent of 245 respondents indicated no validity study activity within the most recent three-year period, 18 percent reported only limited studies (i.e., studies involving only a few departments or programs), and 11 percent reported "unknown" in response to the question about validity studies.

To be sure, ad hoc validity studies involving GRE scores in relation to various measures of student performance or success in graduate study have been conducted from time to time in a variety of graduate school, departmental, and/or disciplinary samples. However, in his review of GRE validation research covering a 20-year period, 1952-1972, Willingham (1973, 1974) could draw upon the results of only 43 studies involving correlations of GRE Aptitude or Advanced Test scores and undergraduate grade point average (UGPA) with diverse criteria of "success," principally the graduate grade point average, but including faculty ratings, departmental examinations, Ph.D. attainment versus nonattainment, and time taken to attain the degree.

These ad hoc studies provided evidence that GRE scores and UGPA were positively related to each of a number of different performance criteria in samples, typically corresponding to graduate departments, from a variety of disciplines. At the same time, most of the studies reviewed were conducted during the 1950's and 1960's, leaving unresolved important questions regarding the correlational validity of these predictors in more recently enrolled cohorts of graduate students. Also, significant questions regarding the predictive validity of GRE scores and undergraduate grades in various subgroups--for example, women, minorities, older students, or foreign students--could not be addressed on the basis of findings of the ad hoc studies reviewed by Willingham.

A Cooperative Studies Rationale

Graduate schools share with all other educational institutions a continuing need for and a responsibility to develop current answers to questions regarding the predictive validity of standard tests and other variables used in screening applicants for admission. These are recurring questions to which schools need up-to-date answers if they are to keep up with changes in student populations, graduate-program characteristics, conditions of test use, as well as changes in the nature of admissions variables per se. To answer these questions, empirical evidence is needed regarding the relationship of admissions variables to clearly defined and relevant, if less than ultimate, performance criteria in representative cohorts, demographic subgroups, and admissions contexts.

Given the complexities of conducting graduate level validity studies, it seems unlikely that concerned graduate schools will be able to monitor GRE predictive validity systematically and thoroughly if they are forced to rely solely on self-initiated validity studies that are based on small departmental samples and are lacking in comparability of design, methodology, sample or cohort definition, and the like.

The Graduate Record Examinations Board in commissioning the Cooperative Validity Studies Project premised its action on the assumption that improvement of GRE validation research is most likely to result from sustained cooperative interaction between all concerned parties: graduate schools and departments, ETS staff, and the Graduate Record Examinations Board.

Cooperative interaction, of course, may take a variety of forms. The Cooperative Validity Studies Project was undertaken to develop and test the effectiveness of one or more validity study models as a basis for implementing GRE validity studies in cooperation with concerned graduate schools and departments.

PART II. DEVELOPING AND IMPLEMENTING COOPERATIVE STUDIES

As indicated above, the Cooperative Studies project was initiated in the fall of 1975, with the general aim of developing and testing models and procedures for facilitating the participation of graduate schools and departments in cooperative GRE validity studies. An immediate objective of the project was to generate up-to-date empirical evidence regarding the correlational validity of GRE Aptitude and Advanced Tests, as well as other preadmissions variables, with respect to relevant criteria of performance in graduate study in clearly defined student cohorts in a variety of departmental settings. It was assumed that experience gained during the course of the project would contribute to the development of longer-term arrangements for regularizing the participation of graduate schools and departments in GRE validity studies.

At the outset, several types of arrangements and/or models were envisaged as potentially useful for developing and implementing cooperative validity studies. For a variety of reasons, it was decided that the most promising approach would be for ETS to develop a structured validity study model, specify the data needed to carry out studies in accordance with the model, and offer to conduct studies and prepare reports for all institutions and departments willing and able to provide the needed data.

It was reasoned that by offering to the graduate school community a sharply focused validity study model with limited data requirements and relatively few conditions for participation, individual graduate schools and/or departments would be able rather quickly (a) to assess the relevance of the model to their interests and circumstances, and, if interested, (b) to indicate their readiness to participate in and provide the data required to complete studies. Findings of studies conducted using this approach, embodying standard data and study design, would be comparable across institutions and departments, and would permit the comparison of findings and the assessment of trends between and across fields of study.

Recruitment of participants was initiated in April, 1976, through a survey of graduate deans of institutions comprising the membership of the Council of Graduate Schools (CGS). The survey was designed primarily to identify prospective participants in cooperative validity studies. However, it was also concerned (a) with ascertaining the types of questions or issues about GRE validity that were current on campus, and (b) with assessing the current status of validation research in CGS schools, especially the extent and nature of validity study activity since 1970.

With a covering letter from the Chairman of the Graduate Record Examinations Board, the survey was mailed in April, 1976, to 344 graduate deans.* A total of 244 deans or their representatives replied. Some degree of interest in the possibility of participating in cooperative studies was indicated by 130 of the respondents. The fact that so many deans reported some degree of departmental interest in the possibility of participating in validation research may be understood best when considered in relation to the extremely low incidence and uneven nature of locally conducted validity study activity reported. Only 38 respondents indicated that studies involving either the GRE Aptitude Test only, or both the Aptitude and Advanced Tests, had been completed since 1970; 30 schools reported limited studies in progress.**

Survey respondents were asked to specify departmental or program areas in which there was an active interest in the validation or further validation of GRE tests as predictors of student performance. It was understood that designation of an area as actively interested would not involve a commitment, but only an indication of readiness to explore actively the possibility of participating in studies, given mutually acceptable models and procedures. As indicated above, 130 respondents (53 percent) indicated some interest in cooperative studies at institutional or departmental levels.

As the next step in the process of recruiting participants, all interested respondents were asked to review and assess the relevance to their interests and circumstances of a short-term validity study model. The salient features of the model are briefly described below.***

- The model focused on the performance of first-time, full-time graduate students who entering a degree program in the fall of 1974, and the fall of 1975. Two entering cohorts were specified in order to augment sample size.

- Departments enrolling 25 or more such students in the two cohorts combined were encouraged to participate by providing in roster format a very limited set of data on each student.

---

* A copy of the covering letter and the survey forms used are included in Appendix B-1.

** A detailed report of survey findings related to the current status of local GRE validity study activity and related issues and concerns, is provided in Appendix B-2. Results are largely consistent with those reported by Burns (1970) which indicated a low volume of validity study activity prior to 1970.

*** A detailed description of the one-year study model is provided in Appendix B-3; a specimen set of data-collection materials is also included.

- More specifically, departments were asked to provide scaled scores on the GRE Aptitude Test (Verbal and Quantitative)* and/or the GRE Advanced Test scores, plus at least one measure of performance (or criterion measure) during the first year of study (e.g., first-year Graduate Grade Point Average or GPA, faculty ratings, end-of-year examinations, etc. Departments were encouraged to provide of an Undergraduate GPA.

- No limit was set on the number of departments for which studies would be made, nor was there any emphasis on specific departments or types of programs. As indicated above, however, the model did involve the explicit delimitation that the samples were to include only first-time graduate students (anywhere). This limitation was included to provide a very necessary measure of control over educational status at point of entry into a program.

- Departments were asked to identify "foreign students for whom English is not the native language" if they were included in a sample, since lack of fluency in English may affect performance on GRE tests. Coding for sex and ethnicity was optional.

- Interested departments could also provide data on other variables of interest--e.g., age at entry, quality of undergraduate school, date of bachelor's degree, etc.

The basic approach was designed to encourage graduate schools and departments to participate in cooperative studies by minimizing the strictures and requirements related to data collection. It was understood that ETS would analyze data and prepare a report of findings for each graduate school, without cost to the participants.

During the period April, 1976, through October, 1976, 44 graduate schools indicated an intention to provide data on one or more departmental samples after reviewing the one-year study model proposed. A total of 35 schools ultimately provided data for one-year studies involving from one to seventeen departments per school.

Several graduate schools with established arrangements for sharing data on admissions-related questions expressed an interest in participating in a study, based on a two-year model, involving the collection of data on first-time, full-time students entering selected departments in the fall of 1974. Four of these schools ultimately provided data for studies designed to assess the predictive validity of GRE scores using two-year cumulative GPA or other criteria.*

---

* These studies were initiated prior to the introduction of the restructured Aptitude Test that yields an Analytical Ability (or GRE-A) score in addition to the Verbal and Quantitative scores.

Thus, 39 graduate schools were recruited as participants in GRE validity studies. These schools are listed in Table 1.

---

* Appendix B-4 provides a brief description of the special studies undertaken in cooperation with these schools. Results of the two-year studies were generally comparable with those that will be described in the subsequent section for the basic one year studies. However, because of differences in definitions and design, results for the two-year studies are not included in the summarizations that are provided in Part III.

Table 1

Graduate Schools Participating in Cooperative Studies

School
Air Force Institute of Technology
Auburn University
Baylor University
Bradley University
Brown University
California State University at Fullerton
Florida Technological University
Fort Hays Kansas State College
Harvard University
Hofstra University
Indiana University at Bloomington
Louisiana State University
Loyola University at Chicago
The Ohio State University
Old Dominion University
Oregon State University
Princeton University
Stanford University
State University of New York at Stony Brook
University of Arizona at Tucson
University of California at Berkeley
University of Colorado at Boulder
University of Hawaii at Honolulu
University of Illinois
University of Kentucky
University of Massachusetts
University of Miami
University of Michigan
University of Missouri at Rolla
University of Montana
University of New Orleans
University of North Carolina at Chapel Hill
University of Notre Dame
University of Oklahoma
University of Texas at Arlington
University of Virginia
University of Wisconsin
Virginia State College
Washington State University at Pullman
Wayne State University

Part III. GENERAL CHARACTERISTICS AND PRINCIPAL FINDINGS OF THE COOPERATIVE STUDIES

During the course of the project, validity studies based on the standard one-year model were completed for 35 graduate schools, and studies based on a special two-year model were completed for four schools. Data were available for from one to 17 departments per school. For each school, data were analyzed by department, and findings were summarized in an institutional report. One institutional report (without identification) is included in Appendix C to illustrate the nature, scope, and limitations of validity studies involving small departmental samples.*

The institutional report provided a basis for organizing and summarizing validity study findings in such a way as to be of greatest direct interest to each participating graduate school. However, since the departmental sample was the basic unit of analysis in all studies, it is more meaningful to examine the general characteristics of samples and data and to summarize findings for departments grouped by field of study or discipline than to do so by institution. Accordingly, in this section, characteristics of the departmental samples, the data employed in the standard one-year studies, and the principal findings of the studies are summarized by field or discipline.

General Characteristics of Samples and Data

The standard one-year studies were designed to assess the relationship of GRE and other predictors, as available, to one or more measures of student performance during the first year of graduate study, in departmental samples from a clearly delimited population, namely, first-time graduate students (anywhere), who were classified as full-time and enrolled in a degree program. Limited data were requested for cohorts entering in the fall of 1974 and the fall of 1975, combined to augment sample size.

The 35 schools participating in one-year studies provided data for over 130 departmental samples meeting study definitions from a wide range of fields or disciplines. The first-year Graduate Grade Point Average (Grad GPA) was provided as the criterion or performance measure for essentially all the samples. Other measures such as faculty ratings or grades in critical courses or course sequences were infrequently provided. Scores on the GRE Aptitude Test (Verbal or GRE-V, and Quantitative or GRE-Q) were also common to all samples. In some samples Aptitude scores were supplemented by an Undergraduate GPA (seldom fully described but characteristically on a scale ranging from A = 4 thorough F = 0) and/or scores on one of the 20 GRE Advanced Tests offered by the GRE Program.

---

* See the report for School A, Appendix C-1.

The departmental samples were typically quite small. Many included fewer than 25 cases and were characterized by missing data patterns--observations or scores on one or more variables were missing for one or more students in most samples.

Some samples included only prospective master's candidates, some only prospective doctoral candidates, and others included both. In essentially all samples that included both prospective master's and prospective doctoral students, first-year departmental programs and evaluation procedures were reportedly comparable for both.*

Table 2 shows for each of 19 fields or clusters of fields (a) the number of departmental samples with data on the respective predictors (GRE Verbal, GRE-quantitative, GRE-Advanced, and/or Undergraduate GPA) and (b) the average (mean) number of students per department with observations on a given predictor. For example, it may be seen that the 22 Bioscience samples had observations on GRE-Aptitude (GRE-Verbal and -Quantitative), but that only 12 of these samples had 25 or more students with Aptitude scores; the mean number of students with Aptitude scores was 26.4. Similarly, only 13 Bioscience samples included observations on the Biology Advanced Test (for an average of 16.8 students) whereas 14 included an Undergraduate GPA, etc.

Several characteristics of the samples available for analysis are clearly discernible in Table 2, including the following:

- As previously noted, GRE Aptitude scores were available for all samples. However, only about half (70 of 138) of the samples included scores on a GRE-Advanced Test; about 58 percent of the samples (80 of 138) provided an Undergraduate GPA.

- The characteristically small size of the departmental samples, which it will be recalled included students in two entering cohorts (fall 1974 and fall 1975, combined), is pointed up clearly in the table. Only 86 of 138 samples included at least 25 students with GRE Aptitude Test scores; only 28 of 70 samples provided data on GRE Advanced Tests for as many as 25 students, and 28 percent of the samples included fewer than 25 students with UGPA as a predictor.

---

* Most of the general features described above are illustrated in the study for School A, Appendix C-1.

Table 2

Number of Samples with Observations on the Respective Predictors and Mean Sample Size Per Predictor

Field/ Department	Number of samples			Mean sample size		
	GRE- Aptitude	GRE- Advanced	UGPA	GRE- Aptitude	GRE- Advanced	UGPA
Biosciences[a]	22 (12)*	13 (2)	14 (5)	26.4	16.8	29.9
Chemistry	12 (6)	7 (5)	8 (7)	32.4	31.3	46.2
Engineering[b]	10 (4)	4 (0)	5 (2)	20.2	10.8	18.2
Mathematics[c]	6 (3)	2 (0)	2 (0)	25.7	17.0	16.0
Physics	5 (3)	4 (3)	2 (2)	36.6**	44.0**	75.5**
Geology; Geophysics	5 (1)	4 (1)	1 (1)	19.6	18.8	39.0
Economics	6 (4)	3 (3)	3 (3)	34.0	36.7	41.7
Anthropology	3 (2)	------	1 (1)	31.7	----	47.0
Education[d]	7 (6)	2 (2)	5 (5)	41.7	29.5	66.4
English	6 (3)	5 (0)	4 (2)	31.7	24.4	36.0
History	10 (10)	7 (3)	8 (8)	34.8	22.8	35.5
Political Sciences[e]	4 (4)	2 (1)	3 (3)	45.3	37.5	52.3
Psychology	12 (10)	7 (5)	7 (4)	43.4	40.0	43.7
Sociology[f]	7 (5)	3 (1)	5 (4)	41.0**	14.3	29.3
Library Sciences	3 (3)	------	3 (3)	39.0	----	39.3
Fine Arts[g]	6 (6)	------	5 (5)	40.7	----	43.8
Music	4 (4)	3 (2)	1 (1)	44.5	28.3	33.0
Philosophy	5 (0)	2 (0)	2 (0)	16.4	8.5	20.0
Languages[h]	5 (1)	2 (0)	2 (0)	17.0	16.5	14.0
Total	138 (86)	70 (28)	81 (56)	32.1	25.0	38.0

*Numbers in parentheses indicate the number of samples for which N = 25 or greater.

**Mean inflated by one relatively large departmental sample.

[a] Includes Oceanography, Marine Environmental Science, Allied Health Science

[b] Includes Engineering and Facilities Management

[c] Includes Computer Sciences, Applied Math and Statistics

[d] Includes Vocational and Adult Education, Educational Administration

[e] Includes Public Administration

[f] Includes Social Work, Urban Planning, Public Policy Studies

[g] Includes Speech and Theater, Drama and Communication, Speech and Communication and Journalism

[h] Includes two Hispanic, one Germanic, one French and one undifferentiated Foreign Languages and Literatures

- Mean sample size, in one or two cases (particularly physics and sociology) inflated by the presence of one atypically large sample, was approximately 32 for the GRE Aptitude analyses, 25 for GRE Advanced Test analyses, and 38 for UGPA analyses.
- The uneven representation of departments across the 19 fields is evident in Table 2. Biosciences were represented by 22 departments, for example, but anthropology by only 3.

Table 3 shows the distributions of the departmental samples according to the degree goals of the students involved. The majority of samples (80 of 138) included both prospective master's and prospective doctoral students, 41 included master's students only, and 15 included doctoral students only. It should be recalled that first-year programs and evaluation procedures were reported to be comparable for both prospective master's and prospective doctoral students in samples including both.*

Coding for sex and ethnicity was optional and quite unevenly available. Several samples, primarily in the physical sciences, included some "foreign students for whom English is not the native language." Only scattered data were available for women, minorities, and foreign students.

Principal Study Findings

The Cooperative Studies were concerned primarily with assessing the relationship of individual predictors, as available, to first-year graduate grade point average (Graduate GPA). As indicated earlier, other performance measures were sometimes provided--faculty ratings, end-of-year examinations, grades in critical courses or course sequences, and the like--but the general Graduate GPA was the "common criterion" in essentially all the studies.**

The correlation coefficient was employed as the index of relationship between a predictor and the GPA criterion. Called a validity coefficient when used to express the relationship between standing on an admissions or predictor variable and standing on a performance or criterion variable, the correlation coefficient is a familiar index that ranges in value from .00 (indicating no relationship at all between two variables) to ± 1.00 (indicating either a perfect positive or a perfect negative association). In studies

---

* Findings for different types of programs are described in a later section (cf., section on subgroup validity).

** For results of one study involving a ratings criterion in addition to Graduate GPA, see Appendix C-1.

Table 3

Distribution of Samples According to Degree Goals of Students Involved

Field/ Department	No. of samples with: Master's only	Master's & doctoral*	Doctoral only
Biosciences[a]	7	14	--
Chemistry	1	10	--
Engineering[b]	5	5	--
Mathematics[c]	1	4	1
Physics	1	2	2
Geology; Geophysics	1	2	2
Economics	-	5	1
Anthropology	-	3	--
Education[d]	5	2	--
English	-	5	1
History	3	5	2
Political Sciences[e]	1	3	--
Psychology	4	3	5
Sociology[f]	4	3	--
Library Sciences	2	1	--
Fine Arts[g]	4	2	--
Music	2	2	--
Philosophy	-	4	1
Languages[h]	-	5	--
Total	41	80	15

*Without significant exception, first-year programs and evaluation procedures were reported to be comparable in these samples which included both prospective master's and prospective doctoral candidates.

[a] Includes Oceanography, Marine Environmental Science, Allied Health Science
[b] Includes Engineering and Facilities Management
[c] Includes Computer Science, Applied Math & Statistics
[d] Includes Vocational and Adult Education, Educational Administration
[e] Includes Public Administration
[f] Includes Social Work, Urban Planning, Public Policy Studies
[g] Includes Speech and Theater, Drama and Communication, Speech and Communication, Journalism
[h] Includes two Hispanic, one Germanic, one French, and one undifferentiated Foreign Languages and Literatures

involving standard admissions measures (such as tests of verbal or quantitative reasoning and an index of past academic performance) and academic criteria (such as first-year grade point average or faculty ratings), obtained validity coefficients are expected to be (and almost always are) positive.*

Collectively, correlational analyses in more than 130 departmental samples yielded a total of over 400 validity coefficients. Results of the departmental analyses grouped by field are summarized in detail in Appendix C-2.** For each departmental sample, Appendix C-2 provides information regarding (a) the obtained validity coefficient for GRE-Verbal, GRE-Quantitative, GRE-Advanced, and/or Undergraduate GPA, as available, with respect to a first-year GPA criterion, (b) means and standard deviations of scores on predictor and criterion variables, and (c) the number of cases used to compute each validity coefficient.

An illustrative summary of findings for departments from one cluster of fields (history, area studies, and anthropology) is provided in Table 4. Certain characteristics of the samples and data alluded to above (e.g., small N's and missing data patterns) are clearly evident in the overall patterns of departmental findings. More importantly, however, the results in Table 4 point up trends that were common to each of the fields or groups of fields considered:

- First, the validity coefficients for GRE and UGPA predictors were overwhelmingly positive, indicating that individuals with higher scores on GRE and UGPA predictors tended to have higher first-year grades;
- Second, this pattern held for samples differing rather markedly in level of GRE Aptitude. Note, for example, that GRE-Verbal validity coefficients were positive in each of 13 samples with mean GRE-Verbal scores that spanned a 200 point range--from a mean of 502 (School 204) to a mean of 698 (School 035).

---

* In a recent summary of the results of over 600 validity studies involving the Law School Admissions Test and Undergraduate GPA as predictors and First Year Average in law school as a criterion, Schrader (1977) lists over 1,200 validity coefficients of which only 11 were negative. For additional discussion of factors involved in evaluating observed validity coefficients in small departmental samples see Appendix C-1, especially pp. 1-8. See also Willingham (1974).

** Appendix C-2 is a summary report, (Tabular Summary of Selected Validity Study Findings, March, 1978) which was prepared for participants in the Cooperative Studies Project.

Table 4

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate Grades: History, Area Studies, Anthropology

School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
School 035 (36,36,17,--)	698 78	603 98	595 54	--	History .27	.39	.39	--	3.56 0.55
School 046 (31,31,13,--)	615 87	535 98	563 60	--	History .10	.20	.05	--	3.81 0.24
School 080 (36,36,--,38)	601 113	541 107	--	3.39 0.42	History .16	-.14	--	.02	3.51 0.35
School 097 (30,30,26,32)	612 87	538 94	563 76	3.56 0.34	History .06	.26	-.10	.40	3.47 0.51
School 103 (34,34,26,26)	594 91	552 121	557 55	3.42 0.38	History .38	.45	.15	.63	3.64 0.31
School 123 (27,27,14,27)	589 87	520 111	534 81	3.14 0.51	History .58	.36	.72	.38	3.47 0.51
School 145 (48,48,--,118)	647 99	546 126	--	3.55 0.35	History .20	.20	--	.20	3.42 0.43
School 145 (37,37,--,55)	651 78	574 103	--	3.32 0.52	Asian Studies .66	.35	--	.12	3.33 0.46
School 221 (43,43,43,32)	649 80	602 107	594 73	3.62 0.28	History .29	.22	.36	.53	3.26 0.44
School 231 (26,26,22,26)	603 109	509 121	570 79	3.34 0.50	History .42	.45	.02	.43	3.46 0.57
School 009 (37,37,--,--)	642 90	582 85	--	--	Anthropology .07	.11	--	--	3.62 0.37
School 145 (39,39,--,47)	649 118	580 109	--	3.62 0.30	Anthropology .41	.30	--	.06	3.58 0.40
School 204 (19,19,--,--)	502 143	483 118	--	--	Anthropology .31	.20	--	--	3.45 0,37

- Third, the mean Graduate GPA (on a scale in which A = 4, B = 3, . . ., F = 0) was higher than 3.4, or approximately "B+," in 11 of the 13 samples, reflecting a consistent tendency for grades to be restricted largely to A's and B's--a pattern that tends to limit the overall "predictability" of differences in grades.

The general tendencies and trends noted above for departmental samples from history, area studies, and anthropology are discernible in study findings for departmental samples from other fields (see Appendix C-2, Tables 1-11). These findings, generally summarized, indicate that:

- First-time graduate students with higher scores on GRE tests (of developed verbal and quantitative reasoning and/or achievement) or with higher undergraduate grade point averages, tended to be better performers, on the average, than their lower-scoring counterparts, when performance was measured by grades earned during the first year of graduate study. This finding held for studies involving a variety of departments from a wide range of disciplines.

Patterns of Validity Coefficients by Field

Results of the departmental analyses clearly support the fundamental premise underlying the use of GRE scores and measures of past academic performance (e.g., an undergraduate average) in assessing the academic qualifications of candidates for admission, namely, that these preadmissions measures should be positively associated with relevant measures of performance in graduate study (such as the first-year graduate grade point average).

However, validity coefficients in small samples (such as those shown in Table 4) have large sampling error. Moreover, in small samples one or two atypical data sets (called "outliers") can have a dramatic influence on both the magnitude and the sign of an obtained coefficient.* Accordingly, validity coefficients obtained in a given departmental sample (a) may not provide reliable information regarding the "true" degree of association between a given predictor and a given criterion--i.e., coefficients will tend to vary substantially from sample to sample in the same department, and (b) do not permit inferences as to the relative validity of two or more predictors.

By aggregating or pooling data for several different departments in a given field (such as history), however, it is possible to arrive at more reliable estimates of validity coefficients for predictors and examine at least tentatively variations in patterns of validity coefficients across different fields of study. One approach to estimating validity coefficients in pooled samples from several departments in the same field involves the use of predictor and criterion variables that have first been departmentally standardized, as decribed below.

---

* See study findings for School A, Appendix C-1, for evidence on this point; see also Appendix D.

Validity coefficients based on standardized variables. Pooling data on GRE scores, UGPA, and Graduate GPA for several departments is complicated by the fact (a) that departments differ in levels of scores on the predictor variables, and (b) that the criterion variable (Graduate GPA) does not have a standard metric from one department to the next. Graduate GPA scales tend to reflect primarily departmental "norms" or standards; accordingly, it does not follow that differences in the average level of grades awarded across departments reflect "real" differences in level of student academic output. A department with "lower" mean GRE scores, for example, may have "lenient" grading standards, and a department with a "higher" GRE mean may have "strict" standards (as suggested, for example, by data for the history departments in Table 4). In any event, it is not possible to generate interpretable validity coefficients simply by combining the original predictor and criterion data for several departments.

However, the problems posed by differences in grading scales and in levels of scores on GRE and other predictors may be dealt with by converting all predictor and criterion variables to a standardized scale within each department, prior to pooling. That is, the GRE scores, UGPA, and Graduate GPA scores for individuals can be expressed as deviations from departmental means in standard deviation units. After standardization, each variable would have a mean of zero and a standard deviation of unity within each department; these standardized scores would be comparable in meaning for individuals without regard to department. Following the departmental standardization of all variables, interpretable correlation coefficients could then be computed based on the standardized variables using data for all individuals from all samples.

The coefficients in Table 5 represent, for each of 19 fields or clusters of fields, predictor-criterion correlation coefficients based on departmentally standardized variables, using data for all departmental samples combined.* The coefficients may be interpreted directly as reflecting the degree of covariation between (a) standing on a designated predictor relative to departmental norms and (b) standing on the Graduate GPA criterion relative to departmental norms. Also shown in Table 5 is the number of cases upon which each coefficient is based. For example, the coefficient of .19 for GRE-V vs. GPA (departmentally standardized) for Biosciences is based on a total of 580 cases from 22 biosciences departments (see Table 2 for the number of samples pooled).

Certain trends are noteworthy, including the following:

- The fields in Table 5 may be thought of as tending to make either primarily quantitative or primarily verbal demands on students. For example, demands on students in biosciences, chemistry, engineering, mathematics, physics, geology, geophysics, and economics may be thought of as more quantitative than verbal, whereas in the remaining fields demands may be thought of as more verbal than quantitative. Inspection of the validity coefficients in the table reveals a tendency for GRE-Q to have higher validity than GRE-V in the quantitative fields (except mathematics), and for GRE-V to have higher validity than GRE-Q in the verbal fields (except psychology and library science, both of which have some quantitative emphases).

In evaluating the coefficients in Table 5, it should be recognized that they are conservative estimates of the degree of covariation between the predictors and "level of academic output" across the entire range of talent represented in the pooled samples--i.e., these pooled coefficients based on standardized variables are lower than those that would be obtained if all the individuals involved were competing in the same department.

---

* One way of illustrating the "meaning" of correlation coefficients of differing magnitude is to show for selected validity coefficients how the relative standing of individuals on a predictor tends to vary with their relative standing on the criterion under consideration, as in the exhibit below (adapted from Schrader, 1971, Table 5.5):

Standing of students on a predictor variable	Expected standing of students on criterion variable (in percent) when:								
	r = .20			r = .30			r = .40		
	Low 20%	Mid 60%	Top 20%	Low 20%	Mid 60%	Top 20%	Low 20%	Mid 60%	Top 20%
Top 20%	13	59	28	10	57	33	7	55	38
Mid 60%	20	60	20	19	62	19	18	64	18
Low 20%	28	59	13	33	57	10	38	55	7

Table 5

Validity Coefficients Estimated in Pooled Departmental Samples Using Departmentally Standardized Variables

Field/ Departments pooled	Size of pooled sample			Validity Coefficients			
	GRE-Aptitude	GRE-Advanced	UGPA	GRE-V	GRE-Q	GRE-Advanced	UGPA
Biosciences[a]	580	219	419	.19	.25	.37	.24
Chemistry	389	219	370	.09	.31	.39	.31
Engineering[b]	202	43	91	.28	.30	.28	.20
Mathematics[c]	154	34	32	.32	.23	.35	.30
Physics	183	176	151	.05	.16	.19	.29
Geology; Geophysics	98	75	39	.05	.06	.11	.37*
Economics	204	110	125	.09	.34	.45	.27
Anthropology	95	---	47	.26	.21	---	.06*
Education[d]	292	59	332	.18	.12	.54	.24
English	190	122	144	.41	.24	.48	.22
History	348	160	284	.31	.26	.21	.30
Political Science[e]	181	75	157	.43	.34	.49	.18
Psychology	521	279	306	.24	.26	.37	.22
Sociology[f]	287	43	146	.43	.30	.54	.55
Library Sciences	117	---	118	.32	.52	---	.33
Fine Arts[g]	244	---	219	.33	.26	---	.31
Music	178	85	33	.24	.11	.21	.23*
Philosophy	82	17	40	.25	.04	.23	.56
Languages[h]	88	33	28	.31	.20	.45	.28
Total	4433	1749	3081	----------Not Computed----------			

NOTE: Validity coefficients are based on departmentally standardized variables. The total number of cases per coefficient is shown under pooled sample size. See Table 2 for the number of departmental samples for which data were pooled.

*Coefficient based on one sample only.

[a] Includes Oceanography, Marine Environmental Science, Allied Health Science
[b] Includes Engineering and Facilities Management
[c] Includes Computer Science, Applied Math and Statistics
[d] Includes Vocational and Adult Education, Educational Administration
[e] Includes Public Administration
[f] Includes Social Work, Urban Planning, Public Policy Studies
[g] Includes Speech and Theater, Drama and Communication, Speech and Communication, Journalism
[h] Includes two Hispanic, one Germanic, one French, and one undifferentiated Foreign Languages and Literature

- In 15 of the 19 fields, one or more GRE predictors tended to have somewhat higher validity than UGPA.

- The overall pattern of coefficients suggests the potential importance of GRE Advanced Tests as predictors of first-year performance in graduate study. Advanced Test scores, the least frequently reported predictor, tended to yield validity coefficients somewhat larger than those for GRE Aptitude or UGPA.

Additional evidence bearing on the relative importance of individual predictors is provided in Table 6, which shows the distribution of sample validity coefficients (based on 10 or more cases) for the respective predictors in departments from "verbal" and "quantitative" fields as defined above.

- The pattern of validities for quantitative fields suggests a primary role for Advanced Tests, followed by Quantitative Aptitude and Undergraduate GPA, while for verbal fields the Verbal Aptitude score supplants the Quantitative score in this pattern.

As will be seen later, the GRE predictors and the Undergraduate GPA each tend to provide some unique information about performance potential.

In evaluating the magnitudes of the coefficients in Tables 5 and 6, it is important to recall that the Graduate GPA criterion frequently was severely restricted in range, being weighted almost always in the direction of "higher grades"--usually only B's and A's. In the circumstances, it is reasonable to infer that observed validity coefficients are somewhat lower than would be the case if differences in student performance were more rigorously and reliably assessed by routine grading procedures.

Comparison with Other Validity Study Findings

Tables 5 and 6 provide summary data indicating the typical levels and patterns, as well as the range, of validity coefficients for GRE and UGPA predictors in samples from a variety of fields or groups of fields. Table 7 relates findings of the Cooperative Studies for selected fields, based on cohorts entering in 1974 and 1975, to findings of studies conducted during the period 1952-72, as summarized by Willingham (1974). A general similarity in the overall patterning of median validity coefficients for the respective predictors is evident for studies that were conducted during two different periods and that involved different samples. The validity coefficients for GRE Advanced Tests, which reflect aptitude, motivation, and substantive achievement in specific fields of study, are typically, somewhat higher than those for other predictors in both periods. This fact provides additional evidence of the potential importance of these tests.

Table 6

Distribution of Validity Coefficients for Groups of "Verbal" and "Quantitative" Fields

Level of validity	Verbal fields*				Quantitative fields**			
	GRE-V	GRE-Q	GRE-Advanced	UGPA	GRE-V	GRE-Q	GRE-Advanced	UGPA
.60 +	2	2	4	3	1	1	1	1
.50 - .59	8	4	5	4	1	6	6	2
.40 - .49	16	8	5	5	4	11	4	5
.30 - .39	10	11	4	10	8	15	4	10
.20 - .29	17	18	4	11	16	17	5	9
.10 - .19	8	15	5	8	10	8	5	4
.00 - .09	4	4	3	4	11	10	2	2
Negative	5	8	1	1	11	5	1	5
Total	70	70	31	46	63	63	29	38
Median Coefficient	.31	.25	.35	.30	.20	.31	.34	.29

NOTE: Includes data for samples of ten or more cases.

*Biosciences, chemistry, engineering, mathematics, physics, geology, economics and other fields (cf., notes to Table 5).

**English, history, sociology, government and political sciences, psychology, education, languages, anthropology and other fields (cf., notes to Table 5).

Table 7

Median Validities Obtained in Cooperative Validity Studies Samples in Selected Fields with Median Validities Obtained in Earlier Validity Studies*

Field(s)	Period covered by studies	Median validity (number of samples)			
		GRE-Verbal	GRE-Quantitative	GRE Advanced	UGPA
Biosciences	1974-1976	.19 (22)	.24 (22)	.37 (10)	.31 (14)
	1952-1972	.18 (7)	.27 (8)	.26 (5)	.13 (2)
Chemistry	1974-1976	.19 (11)	.37 (11)	.41 (5)	.33 (7)
	1952-1972	.22 (14)	.28 (13)	.39 (9)	.27 (7)
Engineering	1974-1976	.26 (9)	.38 (9)	—	.14 (4)
	1952-1972	.29 (11)	.31 (10)	.44 (7)	.18 (4)
Mathematics	1974-1976	.30 (6)	.29 (6)	.40 (2)	—
	1952-1972	.30 (6)	.27 (6)	.44 (5)	.19 (4)
Psychology	1974-1976	.18 (13)	.19 (13)	.32 (8)	.20 (8)
	1952-1972	.19 (23)	.23 (22)	.24 (17)	.16 (15)
Education	1974-1976	.16 (7)	.20 (7)	.53 (2)	.30 (4)
	1952-1972	.36 (15)	.28 (14)	.24 (6)	.30 (5)
English	1974-1976	.30 (7)	.18 (7)	.40 (6)	.27 (4)
	1952-1972	.21 (6)	.06 (6)	.43 (3)	.22 (4)
"Verbal" fields	1974-1976**	.31 (70)	.25 (70)	.35 (31)	.30 (46)
Social science	1952-1972	.32 (11)	.32 (10)	.46 (5)	.37 (5)
"Quant." fields	1974-1976**	.20 (63)	.31 (63)	.34 (29)	.29 (38)

---

*Source of data for earlier studies is a summary by Willingham (1974) of studies during the period 1952-1972. Medians for 1974-1976 are from the Cooperative Studies, using a Graduate GPA criterion. Earlier validities are primarily from studies using Graduate GPA but other criteria were involved in some cases. Number of samples on which medians are based is shown in parentheses.

**Cf., Table 6 and related discussion.

Comparison of Validities at Graduate and Undergraduate Levels

Evidence regarding typical levels of validity coefficients obtained in the Cooperative Studies may usefully be compared with evidence from undergraduate validity studies that have employed a GRE-comparable measure (namely, the College Board's Scholastic Aptitude Test, which yields a verbal score and a mathematical reasoning score) and a measure of previous academic performance (high school GPA or rank in class) versus a first-year grade point average criterion. Such a comparison is provided in Table 8. Several points are important:

- Results of the college-freshman level studies are summarized so as to indicate how validity coefficients tend to be lower in samples that are highly selected on verbal ability than in samples that are more representative with respect to verbal ability.

- Median validities for GRE-Verbal in primarily verbal fields and for GRE-Quantitative in primarily quantitative fields in samples of first-time graduate students (a) are equal to or higher than median validities for comparable undergraduate predictors in samples of college freshmen that are relatively homogeneous with respect to verbal aptitude, and (b) are not markedly lower than validities obtained in more representative college freshman contexts (e.g., colleges using the College Board SAT).

- The most noticeable difference between undergraduate and graduate-level findings is with respect to the validity of the record of previous academic performance: median validity for Undergraduate GPA is rather markedly lower than the median validities for High School GPA or Rank. Graduate students generally may tend to be relatively more highly selected on academic drive and motivation (which undergraduate grades reflect in considerable measure) than are college freshmen generally.

In any event, the findings in Table 8 suggest that despite the recognized limitations of first-year graduate grades (narrow range, over-representation of "higher marks, etc.), when they are employed as a measure of performance, validities obtained for GRE Aptitude tests are similar to those obtained for comparable tests versus college freshman GPA in many undergraduate samples, especially those that are relatively homogeneous with respect to verbal ability. The validity of UGPA for predicting first-year graduate grades appears to be considerably lower than that of high school GPA or rank for predicting first-year undergraduate grades.

Table 8

Median Validity Coefficients and Range of Coefficients in Studies of
Comparable Predictors and First-Year Grade Average
Criteria in Graduate and Undergraduate Settings

Predictors as appropriate to level of study*	GRE studies in graduate school settings	College Board SAT validity studies in undergraduate samples which were:		
		High & homogeneous on SAT-V**	Homogeneous on SAT-V***	Representative SAT-V scores
	Median	Median (range)#	Median (range)	Median (range)
GRE-V (verbal fields)	.31 (70)##			
SAT-Verbal (undergraduate)	------------------------>	.22 (.11 to .44)	.31 (.15 to .46)	.39 (.26 to .54)
GRE-Q (quantitative fields)	.31 (70)			
SAT-Math (undergraduate)	------------------------>	.24 (-.01 to .46)	.27 (.11 to .40)	.33 (.20 to .48)
GRE-Advanced (graduate) (appropriate to field)	.34 (70)			
Undergraduate GPA	.29 (84)			
High School Record	------------------------>	.40 (.32 to .57)	.44 (.26 to .59)	.55 (.33 to .67)

NOTE: GRE validity data are from the current Cooperative Studies. Undergraduate validity data are from Schrader (1971).

* The coefficients in column 1 of the table reflect validity of GRE-Verbal in verbal fields, GRE-Quantitative in quantitative fields, GRE-Advanced Tests as appropriate to a field, and Undergraudate GPA without regard to field. The remaining coefficients are for the Scholastic Aptitude Test (Verbal and Mathematical) and the high school record (either GPA or rank-in-class) in college freshman samples.

**Studies in 18 undergraduate samples having an SAT-Verbal mean above 600 and a standard deviation of 65 or less (college freshman level).

***Studies in 95 samples of undergraduate freshman men and women having SAT-Verbal standard deviations of less than 75. Median values reported separately for men and women by Schrader have been averaged for presentation in this table.

# Approximately 80 percent of the obtained coefficients were within the range specified.

##Number of coefficients (samples) upon which each median is based.

Validity for Subgroups

In the samples submitted for the GRE Cooperative Validity Studies Project, it was not feasible to address systemetically and rigorously questions regarding the comparative validity of GRE tests for important subgroups such as minority students, women, foreign students, or students classified according to type of degree program (e.g., terminal master's, master's-Ph.D. sequence, or doctorate only). The inability to address these questions was due primarily to small numbers or the failure of departments to identify subgroup membership. However, some scattered analyses with very small numbers of cases were carried out. Only limited conclusions can be drawn from these findings, which are summarized generally in this section.

Minority students. Some useful albeit limited evidence bearing on the validity of GRE scores and undergraduate grades for undifferentiated "minority" students (i.e., considering Black, Chicano, Puerto Rican, etc., together) is provided by findings in several very small samples from six graduate schools, summarized in Table 9. In samples from 15 different departments, the number of minority students with predictor (GRE Aptitude) and criterion (Graduate GPA) data ranged from three to 20 students per department. Needless to say, given the small Ns involved, it is important to look primarily for trends, or consistencies that are discernible in the data. It is evident, for example, that validity coefficients tend to be positive in minority, nonminority, and pooled samples.

In samples for School 221, separate analysis was not made of data for nonminority students only. However, in samples from Schools 097, 231, 132, and 145 data were analyzed separately for minority, nonminority, and pooled minority-nonminority samples. In almost every case, it may be seen that validity coefficients for GRE predictors were larger in the pooled sample than in the sample of nonminority students only.

This important finding reflects the fact that minority students typically had substantially lower GRE scores and tended to earn lower grades during the first year of study than their nonminority classmates.

- Figures 1 and 2 provide graphic evidence of this phenomenon. These figures show plots of GRE-Verbal scores versus Graduate GPA for minority students and small random samples of nonminority students in journalism (School 231) and psychology (School 097), respectively. Note that the points or other symbols representing predictor-criterion scores for minority students tend to be clustered in the lower left quadrant of each figure, indicating "below average" GRE-Verbal scores and "below average" Graduate GPA.

Table 9

GRE Aptitude Validity Coefficients in Samples of Minority Students, Nonminority Students and Pooled Minority/Nonminority Students

School/ Field or Department	Minority only		Nonminority only		Pooled	
	V	Q	V	Q	V	Q
School 097						
Psychology * (8, 68, 76)*	19	02	00	-01	14	11
School 132						
Psychology (10, 43, 53)	77	82	23	29	52	54
School 145						
Psychology (20, 69, 89)	33	33	19	17	35	35
Education (10, 40, 50)	26	31	27	09	33	22
School 231						
Journalism (07,25,40)	-36	-11	08	26	21	34
School 204						
Applied Math (07, 25, 40**)	28	52	27	07	27	27
Spanish (06, --, 08)	61	00	--	--	37	-14
Music (04, 58, 66#)	-83	08	17	-08	14	-04
School 221 ##						
Chemistry (03, --, 52)	-34	99	--	--	-21	42
Psychology (09, --, 45)	-27	-38	--	--	42	45
History (04, --, 46)	72	12	--	--	29	22
English (04, --, 55)	67	87	--	--	44	34
Library Sci (04, --, 40)	81	72	--	--	47	59
Hispanic Lang (06, 14)	51	86	--	--	55	70
Public Admin (14, --, 41)	47	17	--	--	32	54

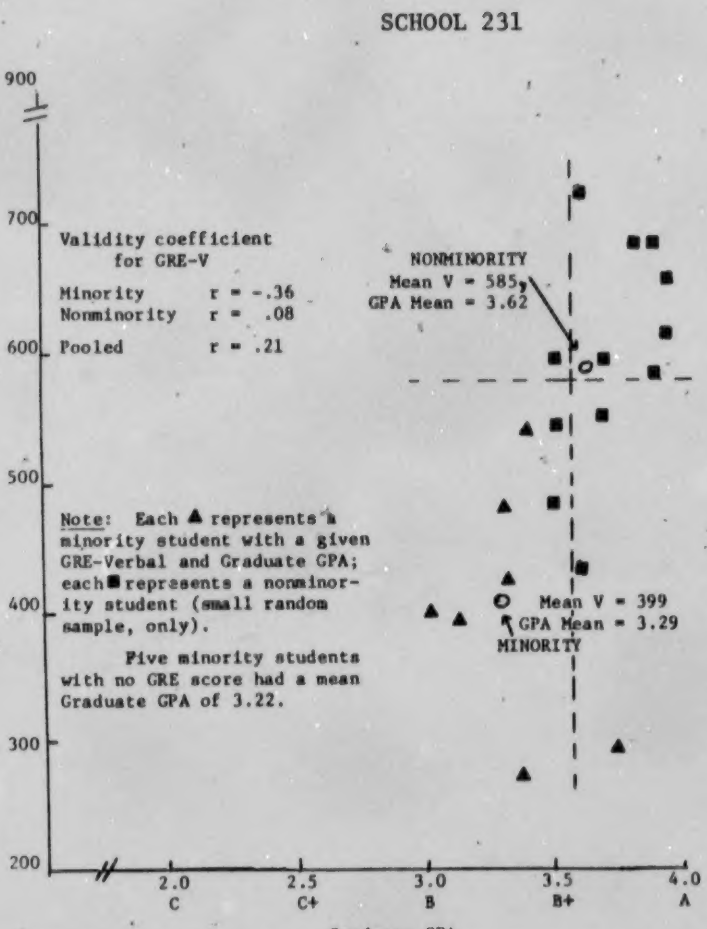
Note: Numbers in body of table, opposite field designations, are correlation coefficients with decimal omitted. The criterion is Graduate GPA.

[*] Numbers in parentheses are Ns used to compute the coefficients. For example, the minority analyses involved 8 cases, the nonminority analyses 68 cases, and the pooled analyses 76 cases in school 097.

[**] Includes eight foreign students.

[#] Includes four foreign students.

[##] Nonminority data were not analyzed separately.



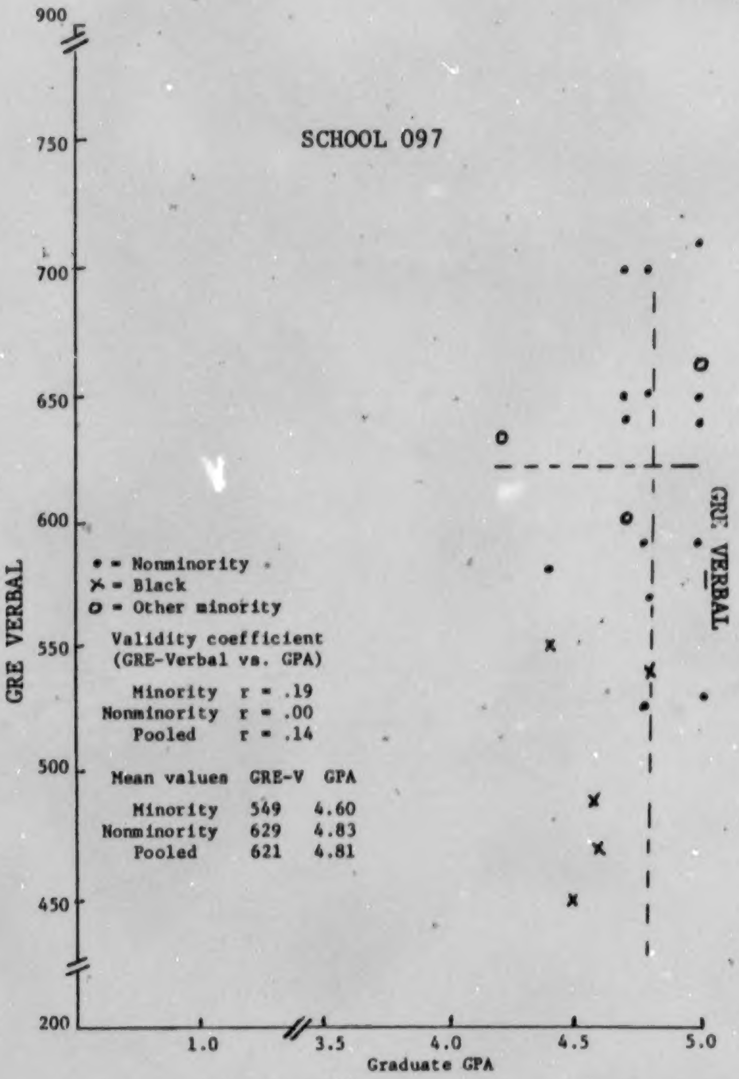
Validity coefficient
for GRE-V

Minority	r = -.36
Nonminority	r = .08
Pooled	r = .21

Note: Each ▲ represents a minority student with a given GRE-Verbal and Graduate GPA; each ■ represents a nonminority student (small random sample, only).

Five minority students with no GRE score had a mean Graduate GPA of 3.22.

Figure 1. Plot of GRE-Verbal scores vs. Graduate GPA for minority and nonminority students in Journalism



• = Nonminority
X = Black
O = Other minority

Validity coefficient
(GRE-Verbal vs. GPA)

Minority	r = .19
Nonminority	r = .00
Pooled	r = .14

Mean values	GRE-V	GPA
Minority	549	4.60
Nonminority	629	4.83
Pooled	621	4.81

Figure 2. Joint distribution of GRE-Verbal scores and Graduate GPA for minority students and a sample of nonminority students in Psychology

GRE VERBAL

This evidence (a) that the GRE validity coefficients tend to be positive for minority students and (b) that minority students with GRE scores well below the average for a given department tend to perform at a level that is also below average for the department, during the first year of study, is consistent with evidence from comparative validity studies in undergraduate and law school settings (see, for example, Linn 1973, 1975; Wilson, 1978).

In evaluating these findings it is important to recognize not only that the findings are based on very small samples in only a few departments, but also that the studies employed only first-year performance criteria. This latter consideration is especially important from the point of view of assessing the validity of preadmissions tests for minority students for whom, it may be argued, first-year performance may not provide a clear indication of performance potential as a result of special problems of transition from undergraduate to graduate study.* Studies employing criteria reflecting performance beyond the first-year in graduate school, as well as additional first-year studies, are needed in order to provide more comprehensive evidence regarding the validity of GRE scores and undergraduate grades for minority students.

Women. Coding for sex, as well as for minority status, was optional. Although the number of women identified was somewhat greater, typically, than the number of minority students, sample size militated against routine analysis and reporting of data by sex. Limited analyses by sex indicate patterns of correlational validity for women that appear to be roughly similar to those for men. Data for several departments in each of two graduate schools are shown, illustratively, in Tables 10 and 11. Median validities for GRE-V, GRE-Q, and UGPA, across departments are summarized below, by sex.

		Median coefficient					
		Men			Women		
School	No. samples	GRE-V	GRE-Q	UGPA	GRE-V	GRE-Q	UGPA
221	(8)	.37	.42	.30	.25	.53	.40
145	(17)	.33	.33	.33	.38	.43	.31

These data, of course, do not permit conclusions regarding the relative validity of GRE and/or UGPA predictors for men and women, respectively, in the two schools involved. However, the observed general trends are consistent with the expectation that GRE scores

---

* The validity of preadmissions tests for predicting the long-term performance of minority and nonminority students has been explored in recent studies at the undergraduate level (e.g., Warren, 1976; Wilson, 1978). These studies suggest that conclusions reached on the basis of comparative validity studies using the first year GPA tend to hold for longer-term cumulative GPA.

Table 10

Correlation of Predictors with Graduate GPA

by Department and by Sex

(School 145)

Department	GRE-Verbal			GRE-Quantitative			Undergraduate GPA		
	F	M	Total	F	M	Total	F	M	Total
Chemistry*	.05	.45	.41	.39	.48	.46	.36	.33	.36
Biology*	.37	.44	.29	.91	.26	.37	.23	.27	.23
Economics*	.13	.23	.17	.49	.33	.37	.64	-.04	.02
Philosophy*	.23	.32	.29	-.86	.06	-.01	.81	.79	.77
Psychology	.49	.25	.35	.57	.29	.35	.21	.13	.17
Anthropology	.16	.57	.41	.13	.44	.30	-.01	.11	.06
Asian Studies	.59	.70	.66	.58	.35	.35	.10	.05	.12
History	.62	.08	.20	.48	.10	.20	.31	.12	.20
English	.17	.35	.23	.34	-.04	.18	.13	.01	.08
Public Policy	.77	.30	.50	.91	.48	.69	.61	.60	.54
Sociology*	.30	.40	.44	.05	.44	.26	.26	.61	.47
Political Science*	.44	.48	.44	.99	.43	.46	-.40	.34	.30
Natural Resources*	.71	.23	.33	.09	.36	.31	.74	.14	.21
Urban Planning	.36	.26	.27	.43	.03	.29	.72	.45	.54
Education	.38	.17	.33	.19	.32	.22	.25	.35	.26
Library Science	.45	.55	.46	.60	.65	.59	.48	.52	.46
Speech	.41	.33	.35	.11	.19	.13	.56	.02	.30
Median	.38	.33	.35	.43	.33	.31	.31	.33	.23

*Coefficients for women based on less than 10 cases. Ns for women ranged from five to 38; for males, Ns ranged from 16 to 51, inclusive.

Table 11

Correlation of Predictors with Graduate GPA

by Department and Sex

(School 221)

Predictor/ group	Chemistry	Psychology	History	English	Library Sci	French	Hispanic Lang	Public Admin	Median validity
GRE-V (M + F)	-.21	.42	.29	.44	.47	.20	.55	.32	(.43)
Female	.27	.14	.24	.57	.46	-.05	.24	.56	(.25)
Male	-.21	.50	.32	.40	.40	.35	.92	.28	(.37)
GRE-Q (M + F)	.17	.45	.22	.34	.59	.43	.70	.24	(.38)
Female	.19	.73	.09	.32	.62	.46	.65	.61	(.53)
Male	.11	.40	.25	.44	.45	.48	.88	.20	(.42)
GRE-Adv (M + F)	.28	.38	.36	.40	n.a.	.36	.57	.50	(.37)
Female	.37	.38	.50	.44		.40	.54	.31	(.39)
Male	.14	.31	.20	.41		.34	.69	.67	(.32)
UGPA (M + F)	.37	.22	.53	.29	.15	.30	.26	.04	(.27)
Female	.46	.11	.69	.43	.34	.38	.22	.68	(.40)
Male	.41	.31	.43	.23	.29	.11	.57	-.07	(.30)
Maximum N	52	45	46	54	40	20	14	40	
Minimum N	14	14	7	17	11	7	5	4	

Note: Advanced Test scores not available in Library Science.

and UGPA should have roughly comparable predictive validity for both men and women. More systematic assessment is contingent upon more representative data.

Foreign Students. In several departmental samples with a relatively high proportion of "foreign students for whom English is not the native language" (as identified by a department), data were analyzed separately for foreign and nonforeign students despite the reduction in sample size entailed by such analysis because of the potentially depressing effect of lack of fluency in English on test performance. Such samples typically were from mathematics, science, or engineering departments in which foreign students exhibited (a) "depressed" GRE-Verbal scores, usually well below the departmental average, but (b) Quantitative and/or Advanced Test scores that tended to be comparable, on the average, with those of their nonforeign classmates.

Results of several limited analyses for foreign and nonforeign samples in chemistry and engineering (see for example, Appendix C-2, Institutional Summary Report, Tables 2 and 3) indicate that patterns of GRE correlational validity were roughly comparable for foreign and nonforeign students. The general pattern of findings suggests that GRE-Quantitative scores are comparable for foreign and nonforeign students entering quantitative fields but that GRE-Verbal scores are not. The evidence provided by the current series of studies is consistent with and in a limited way extends evidence from studies of the performance of foreign students on the GRE Aptitude Test (Harvey & Lannholm, 1961) and of the relationship of GRE Aptitude Test scores to first-year graduate grades in four graduate schools (Harvey & Pitcher, 1963), and in a sample from 24 graduate schools (Sharon, 1971).

In essence, it would appear that "depressed" GRE-Verbal scores of foreign students for whom English is not the native language do not reflect accurately their performance potential relative to nonforeign students in quantitative fields, although among foreign students differences in GRE-Verbal Aptitude tend to be positively associated with differences in graduate grades (e.g., Sharon, 1971).

Questions regarding the comparative performance of and the validity of GRE Aptitude and other tests for foreign and nonforeign students in primarily verbal fields do not appear to have been addressed systematically.

Degree-level. Among the departmental samples involved in the basic one-year validity studies, 37 included prospective master's candidates only, 76 included both prospective master's and prospective doctoral candidates, and only 13 (six from one institution) included prospective doctoral candidates only.*

- o With isolated exceptions in departmental samples that included both master's and doctoral students, first-year programs and evaluation procedures were reported by the respective departments to be comparable for both groups of candidates. In situations in which a given first-year criterion (such as Graduate GPA) reflects differences in performance based on a comparable set of tasks, knowledge of the degree-orientation of the students involved is not essential to the orderly interpretation of validity coefficients.

In the general summary of validity coefficients by department, the degree orientation of students in each sample as reported by the department is indicated (see Appendix C-2, Summary Report, Table 1-11). With the following departmental designations as a basis for classification, selected median validity coefficients were determined for "master's only," "master's and doctorate," and "doctorate only" samples, as follows:

a) median validity of GRE-Q versus GPA in 58 samples from primarily "quantitative" fields (cf., Table 6), and

b) median validity of GRE-V versus GPA in 68 samples from primarily "verbal" fields (cf., Tables 6).

The results, shown in Table 12, indicate that predictor standard deviations and validities tended to be lower in doctoral samples than in either master's samples or samples that included both master's and doctoral students.

---

* In explaining the very small number of "doctorate only" samples, it is important to recognize that the study, by design, was restricted to first-time graduate students only. Many departments rely heavily on the recruitment of master's degree holders to obtain their prospective Ph.D. candidates. One departmental chairman, commenting on the small number of cases for which he could supply data, expressed surprise at "discovering" how few of his doctoral students had begun their graduate work in the department. The Cooperative Validity Studies Project was concerned with a clearly defined population that did not include students admitted to Ph.D. programs after having earned a master's degree or equivalent. It should not be assumed that work completed by such individuals during their first year in a program is comparable with that completed by first-time graduate students during their first year in a program. Validity study models for master's-holders entering doctoral level programs will be needed in order to deal with this general set of circumstances.

It is important to keep in mind in evaluating these results that a very small number of doctorate-only samples is involved and that several of these were from only one graduate school; consequently, specific detail should not be considered significant. Nonetheless, the observed pattern is of interest, because it suggests that the lower median validities for the "doctorate only" samples in this series of studies may be due primarily to greater restriction of range on the respective predictors in these samples. As indicated in Table 13, median validity tends to decrease as predictor standard deviation decreases without regard to degree-level of samples. More representative data will be needed to determine how general this pattern may be.

Summary: subgroup analyses. All findings with respect to "subgroup validity" should be viewed as suggestive only, and as incidental to the primary objectives of the Cooperative Validity Studies Project. Undue emphasis should not be placed on specific detail in evaluating the findings. Results for very small samples suggest that in graduate school, as in other academic settings, standard test scores are positively associated with grade point average for minority as well as for nonminority students and that lower-than-average test scores for minority students presage lower-than-average first-year grades.

- In contrast, limited analysis suggests that lower than average GRE-Verbal scores for foreign students (for whom English is not the native language) in heavily quantitative fields probably do not consistently presage lower than average first-year performance for foreign students relative to the departmental average. The first-year GPA level for foreign students appears to be roughly consistent with their average level on GRE-Quantitative Aptitude.

- Patterns of validity for women and men, respectively, appear to be roughly comparable, as expected.

- And finally, trends observed in connection with the analysis of validities for degree-level subgroups are consistent with familiar restriction-of-range axioms (cf., Appendix C-1, pp. 3-6; also Table 8 and related discussion).

However, in none of the subgroup analyses involving women, minority, or foreign students could systematic attention be given to the many complex questions that are involved in the rigorous determination of the comparative validity and "fairness" of preadmissions measures for the respective subgroups (Linn, 1973).

It is important to recognize that building a reliable body of empirical evidence bearing on subgroup validity will require the participation of graduate schools and departments in cooperative validity studies designed especially to collect data on the subgroups of interest.

Table 12

GRE Median Validities and Standard Deviations for Samples Classified According to Degree Orientation of Students

Type of sample	Verbal fields*			Quantitative fields**		
	No. of dept.	GRE-V S.D. (median)	GRE-V validity (median)	No. of dept.	GRE-Q S.D. (median)	GRE-Q validity (median)
Master's only	22	94	.30	15	85	.25
Master's & doctor's	38	97	.32	38	82	.34
Doctor's only	8	80	.25	5	65	.00
All samples	68	95	.31	58	82	.31

*Biosciences, chemistry, engineering, mathematics, physics, geology and geophysics, etc. (cf., Table 6 and related discussion).

**English, history, psychology, languages, government, sociology, education and other fields (cf., Table 6 and related discussion).

Table 13

Median GRE Validities for Samples Classified According to Size of GRE Standard Deviation

Aptitude standard deviation	Quantitative fields		Verbal fields	
	No. of GRE-Q S.D.	GRE-Q validity (median)	No. of GRE-V S.D.	GRE-V validity (median)
100 +	12	.30	27	.35
80 - 99	20	.37	31	.30
60 - 79	20	.25	12	.20
Below 60	6	.15	0	---
Total	58	.31	70	.31

NOTE: Grouping of fields as for Table 12.

PART IV. THE PROBLEM OF COMBINING PREDICTORS IN SMALL GRADUATE DEPARTMENTS

In validation research generally, it has been found that an appropriately weighted composite of scores on standard admissions tests and an index of past academic performance normally yields a higher validity coefficient (coefficient of multiple correlation) than either test scores or the past academic record considered separately. In large-sample studies, questions regarding the most appropriate weighting of admissions variables for the purpose of predicting a specified criterion are addressed directly by using appropriate multivariate methods, principally multiple regression analysis.

Given a criterion such as first-year GPA and scores on several predictors, the basic output of validity studies employing multiple regression analysis includes, in addition to validity coefficients for each of the predictors considered separately,

1. a multiple correlation coefficient reflecting the relationship of all the admissions variables, considered jointly, to the criterion;

2. standardized regression weights (called beta weights) indicating the contribution of each admissions variable in an optimally weighted composite-predictor; and

3. a regression equation specifying the (multiplier) weights to be applied to the scores on the admissions variables in order to obtain a composite-predictor score that is optimally-weighted for predicting the criterion variable under consideration.

In large samples, the multiple regression model provides a systematic basis for determining how much each of several admissions variables contributes to the overall effectiveness of prediction, and the multiple regression equation summarizes scores on several admissions variables by giving to each score a weight that reflects its unique contribution to an optimally weighted composite predictor.

In practice, regression weights are estimated in a given sample, such as one year's entering class, and applied in summarizing the admissions scores of candidates for admission to subsequent classes. Even in larger samples, there are questions regarding the stability of the estimated regression weights (e.g., Wainer 1976, 1978). However, especially in small samples the results of a regression analysis tend to reflect too closely possibly idiosyncratic patterns of interrelationships in the sample data ("overfitting" the data), and thus do not provide reliable estimates of the "true" or population weights. Sampling error for observed regression weights in small samples is great, and weights developed in successive small samples of first-year students will tend to fluctuate widely.

Because of the consistently small size of the departmental samples that are involved in graduate-level validity studies, it is not feasible to use multiple regression analysis routinely in order to determine "optimal weights" for available predictors (e.g., GRE-V, GRE-Q, and Undergraduate GPA or UGPA) that could be used by a department to form a locally relevant composite predictor. There is every reason to believe, however, that a combination of GRE and UGPA should lead to improved prediction of a given criterion. Willingham (1974), for example, reported a median multiple correlation of .45 for 24 departmental samples for a combination of GRE-V, GRE-Q, and UGPA, optimally weighted for predicting first-year Graduate GPA, whereas median validities for these predictors considered separately were in the .20 to .30 range. In the Cooperative Studies, multiple correlation coefficients were reported illustratively for some departmental analyses.* In 27 samples from a variety of fields, the median multiple correlation for the same set of predictors with respect to the GPA criterion was .43, as compared with median validities in the .30 to .35 range for the predictors considered separately.

In view of the potential benefit (improvement in predictive validity) likely to accrue from combining predictors, the problem of determining wieghts for GRE and UGPA variables that a given department might use to form a composite predictor is an important one. In small departmental samples (and other situations in which there are insufficient data to provide reliable estimates of weights for commonly used predictors), there is reason to believe that workable solutions to the problems involved in combining predictors may be found in approaches involving pooling data for several departments within the same field. The basic rationale underlying approaches involving pooling data for several small samples in a given field (say, chemistry) is that there are substantial elements of similarity in the general types of tasks performed by students from department to department within the same field.*

---

* Because of small sample size, multiple regression analysis was not employed routinely in the Cooperative Studies, and multiple correlation coefficients were reported, in selected studies, primarily to facilitate discussion of the principles and problems involved in developing and evaluating the predictive value of weighted composites of predictors (cf., Appendix C-1 pp. 7-8, and 16-17; see also Appendix C-2, Tables 1-11).

Testing a "Common Weights" Hypothesis

Given data consisting of a common set of predictors and a comparable criterion variable for each of several departmental samples in a given field of study (say, chemistry), it is reasonable to ask whether (regression) weights for predictors as estimated from individual departmental data differ significantly from weights that may be estimated by the use of pooled data for all the departments involved.**

---

* In a GRE-Board sponsored study, Boldt (1975; cf., Appendix A) called attention to the importance of assuming significant elements of similarity in the types of activities involved in educational or occupational pursuits of the same kind that are being carried out in different locations. Even though the tasks involved in a first-year chemistry program, for example, may be conducted at different levels of difficulty and with differing emphases from department to department, it is reasonable to assume that the general underlying similarities are at least as great as the differences in tasks.

** In graduate level validation research, and in other validity study settings as well, questions may be raised regarding the practical utility and relevance for admissions decisions of gains, if any, that may accrue from differential weighting of a common set of predictors for each of several different but similar prediction contexts. It may be argued that the level of precision implied by "unique weighting" is not justified considering (a) the presumed multidimensional nature of the assessment process that culminates in admissions decisions, (b) the fluctuations that occur from year to year and sample to sample in the magnitudes--even the signs--of weights generated by within-groups analyses, (c) the limitations of performance criteria employed in studies, lack of a working consensus regarding one criterion as being the most appropriate or representative, and the presumption of a positive correlational manifold among all potential criterion variables, (d) the probably high degree of similarity across "similar" settings in the relative demands placed on general verbal and quantitative abilities, and (e) the typically high correlation between "reasonably" weighted composites of predictors. Several liberal arts colleges, law schools, or graduate chemistry departments, respectively, are likely to have at least as many elements of similarity as of difference in their patterns of demands upon student verbal and/or quantitative abilities. It is not unreasonable to hypothesize that a limited number of sets of weights for a common set of predictors should be sufficient for purposes of within-group prediction in identifiable clusters of similar selection settings. See Wainer (1976, 1978) for an examination of the "weight-fluctuation" problem and rationales for dealing with it.

Evidence that predictor weights estimated from individual departmental data do not differ significantly from weights estimated from data pooled from all departments would support that important underlying assumption regarding general similarities across departments in the same field. In addition, it would suggest that a single "solution" to the problem of combining predictors might be applicable in each of several departments in a given field. One important practical implication is that each of the departments could use the pooled-department weights in developing a composite-predictor with local relevance, validity, and utility.

Data and Analytic Approach*

Data for 54 departmental samples from five fields were selected for the exploratory analysis: biosciences including departments of zoology, botany, forestry, natural resources, marine science, general biology, chemistry, psychology, English, and history. The samples were from 25 different graduate schools. Table 14 shows for each of the five fields the number of samples involved in the analyses, and the median and range of the sample sizes.

Note in Table 14 that there were fewer samples with GRE-Verbal, GRE-Quantitative, and UGPA (V,Q,U), than with Verbal and Quantitative (V,Q) only. Since scores on the GRE Aptitude Test (Verbal or V, Quantitative or Q) constitute the most commonly available set of predictors, it was considered desirable to test the common weights hypothesis for V,Q as the independent variables (called the V,Q analysis), and then conduct a second series of analyses involving V,Q, and UGPA (called the V,Q,U analysis) by using data for a reduced number of samples.

The analytical approach employed in testing the common weights hypothesis is outlined below, assuming the availability of three predictors (V,Q, and UGPA) and a "common" criterion, namely, Graduate GPA for several departmental samples. (It is important to note that the first step in the analytic procedure described below is to standardize graduate GPA within each department prior to pooling data in order to control for differences in the grading scales):

Let us consider only one type of department at a time--say, chemistry. For the ith school let Y_i denote the graduate GPA in that school. Y_i is a variable defined on all students in the given department in school i.

---

* The consultative assistance of Paul Holland, office of Data Analysis and Research at ETS, who suggested the analytic approach used in testing the common-weights hypothesis, is acknowledged with appreciation.

Table 14

Number of Samples (Departments) in V,Q and V,Q,U Analysis, and Data on Size of the Samples

No. of departments per analysis	Biosciences	Chemistry	Psychology	English	History
No. of V,Q samples*	(19)	(9)	(12)	(6)	(8)
Median N	25	29	40	36	30
Range of N's	(6 - 43)	(10 - 93)	(20 - 89)	(19 - 54)	(25 - 48)
No. of V,Q,U samples**	(13)	(6)	(8)	(5)	(7)
Median N	28	19	38	34	29
Range of N's	(6 - 43)	(11 - 92)	(22 - 89)	(14 - 51)	(25 - 48)

*Number of samples in analyses involving Verbal and Quantitative scores (V,Q) as independent variables.

**Number of samples in analyses involving Verbal, Quantitative, and UGPA (V,Q,U) as independent variables.

1) Standardize Y_i to have mean 0 and variance 1 within school i. This is done because the schools may have different grading systems we are pooling them in a regression so we want to remove this source of between-school differences, at least superficially.

2) Using the data from all students with complete records from all schools of the given department type, estimate equations of the form:

$$(1) \qquad \hat{Y}_i = \hat{a}_i + \hat{b}_1 V + \hat{b}_2 Q + \hat{b}_3 \mathrm{UGPA} \qquad i = 1,\ldots,n.$$

Note that this estimates common weights for V, Q, and UGPA across all schools but allows each school to have a separate intercept term, a_i.

3) Now the question naturally arises: are there sufficient data from the ith school to determine that it has weights b_{1i}, b_{2i}, or b_{3i} that are different from the pooled weights? This is done for b_{1i} by fitting equations of the form:

$$(2) \qquad \begin{aligned} \hat{Y}_i &= \hat{a}_i + \hat{b}_{1i} V + \hat{b}_2 Q + \hat{b}_3 \mathrm{UGPA} \\ \hat{Y}_j &= \hat{a}_j + \hat{b}_1 V + \hat{b}_2 Q + \hat{b}_3 \mathrm{UGPA} \text{ for } j \neq i. \end{aligned}$$

The actual fitting of these equations will be done by least squares and will use indicator variables and their products with V,Q and UGPA to fit equations like (1) and (2).

4) The test for whether or not a separate weight is needed for V in school i is the 1-degree-of-freedom F-test obtained by comparing the residual sums of squares from (1) and (2) in the usual way.

Regression Results When Data Were Pooled

Following the foregoing analytical approach, pooled departmental data were used to estimate regression weights and multiple correlation coefficients (a) for V and Q in one series of analyses, and (b) for V,Q, and Undergraduate GPA in a second. Table 15 summarizes the pooled within-department regression results for the V,Q and V,Q,U analyses, respectively, for each of the five fields. The weights shown represent estimates of weights for standardized predictor scores, and the multiple correlation yielded by the combined predictors based on all the available data. Several features of the findings are noteworthy, including the following:

Table 15

Results of Pooled Departmental Data Regression Analyses, by Field

Field	No. of samples pooled	Total no. of cases	Standard regression weight			Multiple correlation coefficient
			GRE-V	GRE-Q	UGPA	
Biosciences	(19)*	458	.177	.206		.292
	(13)**	390	.178	.240	.208	.390
Chemistry	(9)	300	-.077	.368		.343
	(6)	203	.005	.289	.330	.444
Psychology	(12)	518	.184	.187		.286
	(8)	326	.234	.178	.200	.386
English	(6)	215	.352	.110		.394
	(5)	151	.368	.084	.183	.437
History	(8)	262	.197	.155		.294
	(7)	228	.155	.148	.307	.415

*Number of samples in analyses involving V and Q only as independent variables.

**Number of samples in analyses involving V, Q, and UGPA as independent variables.

- Note that combining GRE and UGPA scores results in increased validity. The multiple correlation coefficients for V, Q, U are considerably greater than those for V,Q only. This result is expected.
- In one field, chemistry, GRE-Verbal tends toward zero-weighting;* in English, Q makes a comparatively small contribution as compared with V, while in the remaining fields V and Q tend toward equal weighting.

Testing deviations of departmental weights from the pooled estimates. Following the procedures outlined in the analytic approach, above, tests were made of differences between predictor weights estimated by using data for individual departments and the predictor weights estimated by using the data for all the departments. Table 16 summarizes the outcomes of tests.

- In the analyses involving V,Q as the independent variables or predictors, departmental weights for V were found to deviate significantly (p. $\gtrless$ .05) from the pooled estimates in only six of 54 samples, across all fields, and deviant departmental weights for Q were indicated in only five of the 54 samples.
- In the V,Q, U analyses, few sample weights were significantly deviant--in only two of 39 samples, the weight for V was deviant(p. $\gtrless$ .05); in only three of 39 samples, weights Q or UGPA differed significantly from the pooled estimates.

---

* In evaluating the negative coefficient for GRE-V in this departmental analysis, it is important to keep in mind that when a negative regression weight is obtained for an academic predictor, the predictor involved can be excluded from the set of predictors involved in the analysis. In this case, all the information of value for estimating first-year grades is being provided by GRE-Q (in the V,Q analyses), or GRE-Q and UGPA in the three-predictor analysis. Moreover, a negative regression weight may be obtained in circumstances in which the predictor involved has a positive validity coefficient when considered separately. Consideration of this phenomenon, known as "suppression effect," is outside the scope of this report. However, it is of considerable importance in graduate-level validation research because it tends to occur under conditions that may be encountered in fields that are either heavily "quantitative" or heavily "verbal" (such as chemistry in this particular analysis) when both verbal and quantitative ability measures that are moderately highly related are included in a prediction battery. [Cf., discussion of the problem involved using a GRE-Aptitude total score, Appendix C-1, pp. 7-8; see also Wilson (1974) for evidence of recurring suppressor effects in undergraduate settings.]

Table 16

Summary of Outcomes of Tests of Differences
Between Sample and Pooled Estimates
of eights for Predictors

Field/ independent variables		GRE-Verbal		GRE-Quantitative		Undergraduate GPA	
		No. of tests	No. of deviant weights*	No. of tests	No. of deviant weights**	No. of tests	No. of deviant weights***
Biosciences	V,Q	(19)	1	(19)	2	Not applicable	
	V,Q,U	(13)	0	(13)	1	(13)	0
Chemistry	V,Q	(9)	2	(9)	0	Not applicable	
	V,Q,U	(6)	0	(6)	0	(6)	0
Psychology	V,Q	(12)	3	(12)	2	Not applicable	
	V,Q,U	(8)	1	(8)	1	(8)	1
English	V,Q	(6)	0	(6)	0	Not applicable	
	V,Q,U	(5)	0	(5)	0	(5)	0
History	V,Q	(8)	0	(8)	1	Not applicable	
	V,Q,U	(7)	1	(7)	1	(7)	2
All fields	V,Q	(54)	6	(54)	5	Not applicable	
	V,Q,U	(39)	2	(39)	3	(39)	3

*In V,Q analyses, the weight for V in a sample, estimated using a common (pooled) estimate for Q, differs significantly from the pooled weight for V ($p \leq .05$); in V,Q,U analyses, weights for Q and UGPA are constant in each test for departmental weight for V.

**In V,Q analyses, the weight for Q in a sample, estimated using a common (pooled) estimate for V, differs significantly from the pooled weight for Q ($p \leq .05$); in V,Q,U analyses, weights for V and UGPA are constant in each test for departmental weight for Q.

***In these analyses, the weight for U in a sample, estimated in an equation involving pooled sample weights for V and Q, differs significantly from the pooled estimate of the weight for U ($p \leq .05$).

- No systematic direction was indicated for the test results. That is, in seven samples in which weights for one or more predictors were identified as deviant, the sample weight was larger than the pooled estimate, while in eight cases the sample weight was smaller (and in some instances, anomalously negative).

Examination of samples with deviant weights. The test results indicate that the data generally conform to the "common weights" hypothesis. However, what about the deviant departments? Are there characteristics in the data that may help to explain the "deviant" outcomes? To shed light on these questions, a detailed examination was made of the data in all departments in which one or more predictors were identified as having regression weights differing significantly from the pooled estimate.

In almost every instance, examination of the original data for the 13 samples involved (with Ns ranging from 11 to 52) revealed conditions that help to account for "deviant" regression weights. Detailed results of the examination are outlined in Appendix D. However, the essential nature of the findings may be summarized as follows:

- In samples characterized by atypically high positive regression weights for GRE-V and/or GRE-Q, the observed result was associated with one or more atypical data sets for individuals who were identifiable in certain ways as "atypical"--e.g., members of minority groups with very low test score(s) and also very low graduate grades. [See detailed departmental analyses in Appendix D.]

- In samples with anomalous negative coefficients, outcomes were clearly associated with one or two extremely atypical data-sets or outliers that heavily influenced results--e.g., one individual with unusually low standing on a predictor and unusually high standing on the criterion, or vice versa.*

---

* Careful examination of the detailed data in Appendix D will reinforce this important point regarding the impact that one or two aberrational data sets, or outliers, can have on the magnitude and/or the sign of validity coefficients in small samples. Negative coefficients, of course, are anomalous--i.e., coefficients reflecting the relationship of academic predictors (such as GRE scores) to academic criteria (such as grades) should be positive, a priori. Given the potential for anomalous outlier impact, the overwhelmingly positive distribution of validity coefficients obtained in the Cooperative Studies in data for very small departmental samples indicates a remarkable degree of underlying regularity in such data. Attention to sample definition, however, clearly is necessary in order to avoid confounding results.

Implications of the Findings

On balance, the findings of these exploratory analyses lend support to the common weights hypothesis and to the important assumptions underlying the use of pooled data for several small departments in GRE validation research. The following points are important:

- Results of the regression analysis, per se, clearly indicate that prediction should be improved by an appropriate combination of available GRE and UGPA predictors.
- It may safely be assumed that small departments are not (and are never likely to be) in a position to obtain reliable estimates of predictor weights by using local data only.
- Results of the series of tests, on balance, support the common weights hypothesis. They enhance the prospect that pooled-data analysis may provide solutions to the problem of combining predictors that can be applied in local departmental settings.
- The findings point up the importance (a) of validity study models employing data that are generally comparable across departments, and (b) of the concurrent participation of several departments from the same field in cooperative validity studies.

It is important to recognize, in connection with the foregoing points, that it is not necessary to hypothesize a strictly "common weights" solution to the problem of estimating predictor weights, using pooled data approaches, that may have local applicability, validity, and relevance for several graduate departments in a field. So-called Bayesian methods of analysis have shown promise as a means of "adjusting" locally derived regression weights on the basis of findings in aggregated or pooled samples (e.g., Boldt, 1975; Rubin 1978). These methods have been applied successfully in contexts involving relatively large "local" samples (e.g., in undergraduate and law school settings). The important consideration is that pooled data approaches that have had demonstrated effectiveness in certain settings appear to offer special promise for graduate-level validation research.

The present exploratory study represents a useful first step. It should be kept in mind that the departmental samples involved are not necessarily representative of all graduate departments from their respective "fields." Further empirical study is needed and appears to be fully warranted on the strength of the findings that have been reviewed.

V. RETROSPECT AND PROSPECT

For a variety of reasons, assessment of the predictive validity of GRE tests in graduate school settings has not been carried out on a regular basis. The volume of validity studies involving GRE tests has been low relative to the number of settings in which validity studies could and should be conducted and as compared with the volume of studies conducted in undergraduate and certain professional school settings where circumstances have been conducive to the widespread, routine application of standard validity study models and procedures.

In commissioning the GRE Cooperative Validity Studies Project, the Graduate Record Examinations Board accepted, at least tentatively, certain basic premises, as follows:

1) In light of past experience, and considering the inherent complexities involved in conducting graduate-level validity studies, it seems unlikely that concerned graduate schools and departments will be able to monitor GRE predictive validity thoroughly and regularly through self-initiated studies alone;

2) The participation of graduate schools and departments in the validity study process on a regular recurring basis, the generation and widespread dissemination of up-to-date and interpretable information regarding GRE predictive validity in a variety of contexts, and the improvement of validity study procedures generally, are goals that are shared by all parties concerned with GRE development and use;

3) Attainment of these shared goals is most likely to be realized through sustained cooperative interaction between all concerned parties, namely, graduate schools, Educational Testing Service, and the Graduate Record Examinations Board.

The Project was charged with developing and exploring the utility of one or more specific models for facilitating and encouraging the participation of graduate schools and departments in GRE validity studies. It was assumed that experience gained during the project would contribute to the development of arrangements and procedures through which the GRE Program might facilitate the regular participation of graduate schools and departments in validity studies.

The activities involved and the findings generated in carrying out this charge, over a three-year period, with the sustained support and encouragement of the Graduate Record Examinations Board, have been described in detail. The graduate school community was invited to participate in cooperative studies based on a sharply focused, highly structured validity study model, with limited, clearly defined procedural and data requirements; ETS offered to conduct studies for and report findings directly to each graduate school willing and able to provide the needed data for one or more departmental samples, at no cost to the participating school. It was reasoned that data generated by this procedure could be compared across departments within institutions and would facilitate the comparison of findings and the assessment of trends within and across fields.

This approach was successful in enlisting the cooperation of 39 graduate schools, represented by from one to 17 departments per institution, in validity studies. Individualized institutional reports were prepared for each participating school. More than 150 data-sets, generally corresponding to departments and representing over 19 fields or clusters of fields, were analyzed. The data generated by these studies permitted the analysis of trends in patterns of correlational validity for GRE Aptitude and Advanced tests and Undergraduate GPA in recently enrolled cohorts of first-time, full-time graduate students, primarily with respect to first-year Graduate GPA criteria. A report summarizing the findings of all institutional studies was sent to each participant in the Project.

The findings indicate that the frequently cited problems of conducting graduate-level validity studies are very real. Problems associated with small samples, unstable weighting, restriction of range, criterion selection, and so on, are inherent in graduate school settings and must be dealt with in all graduate-level validity studies. However, experience during the Project indicates quite clearly that it is possible to conduct basic validity studies yielding useful, interpretable results despite these problems.

For analyses involving very small departmental samples to yield useful results, it is important to make sure that the samples are clearly defined and relatively homogeneous with respect to student educational status at entry, and that students are engaged in comparable pursuits over a defined study period. It is believed that careful attention to the problem of sample definition contributed significantly to the generally interpretable nature of the findings obtained in the Cooperative Studies. Only first-time, full-time, degree-seeking students were included in the samples (and the findings, of course, apply only to such samples).

Small sample results become increasingly meaningful as data from several departments in each of a variety of fields can be aggregated to provide normative perspective for assessing trends in patterns and levels of correlational validity for several predictors within and across fields.

Results of special analyses indicate that pooling procedures, using data for comparable sets of predictor and criterion variables for several small samples in the same field, have considerable promise as a basis for arriving at reliable estimates of validity coefficients and weights for predictors--weights that may be used by small graduate departments in a given field to combine available predictors in such a way as to form a composite-predictor having local relevance and predictive validity.

The findings of the Project provide firm support for the interpretive rationale posited at its inception, namely, that measures of developed ability and achievement (such as GRE scores and Undergraduate GPA) should tend to be positively related to measures of performance in graduate study (such as the Graduate GPA). The overwhelmingly positive pattern of relationships found in this project add to evidence from earlier studies that GRE scores and undergraduate grades provide relevant information that can be useful as part of the complex process of screening applicants for admission to graduate school. Limited evidence was also provided regarding the validity of GRE tests for women, minorities, and foreign students. However, analyses were based on very small samples. Special efforts will be needed in order to obtain more comprehensive validity data for these and other subgroups of special interest (e.g., older students, part-time students).

It is believed that the results attained during the Cooperative Studies Project indicate the validity of the premises underlying commission of the project by the GRE Board. All parties to GRE development and use have a responsibility to develop current answers to questions regarding the predictive validity of GRE and other admissions variables--answers calling for empirical evidence regarding the relationship of these variables to clearly defined and relevant criteria of performance in representative cohorts, demographic subgroups, and graduate admissions settings. Answers to these questions must be kept current to monitor changes in validity that may occur with changes in student populations, graduate programs and curricula, grading standards, conditions of test use, and the characteristics of the GRE tests themselves.

Procedural and other arrangements are needed to facilitate the recurring participation of all GRE-using graduate schools and departments in basic, standard validity studies. Recurring participation in studies is even more important in graduate school settings, with characteristically small departmental samples as the units of analysis, than in undergraduate and law school settings with large entering cohorts as the units of analysis. In these latter settings, the availability of program-supported admissions-related research and validity study services has been directly responsible for the development of regular patterns of institutional participation in validation research.

Sustained cooperative arrangements involving the GRE Board, Educational Testing Service, and concerned graduate schools can contribute directly to the regularization of the graduate-level validity study process. Several features of existing program-related validity study services that are likely to be relevant to the development of long-term GRE Program validity study services are as follows:

1. Primary focus on a clearly defined study period (typically, the first year of study), and a basic, limited core of validity study data (typically, test scores, a measure of undergraduate performance, and a criterion measure), with some options for extending standard studies on an ad hoc basis;

2. Maximum use of the program's central data file to facilitate the collection of validity study data and the extension of services (descriptive statistics etc.) to test-users;

3. Regular, publicized cycles of participation;

4. No cost to participating institutions for analysis and reporting, with funding on a programmatic, continuing basis.

Plans for a continuing GRE Validity Study Service embodying features similar to those outlined above have been developed by the GRE Program staff and approved by the GRE Board. The implementation of such a service in the face of the complexities characteristic of graduate school organization will not be easy. However, its development offers an exciting and challenging opportunity for continued collaboration among the GRE Board, Educational Testing Service, and concerned graduate schools, aimed at regularizing the GRE validity study process. Such regularization is a necessary step toward the goal of assuring that those who make critical decisions to accept some and reject other applicants for graduate study can be guided by up-to-date and interpretable information about the implications of GRE scores for those decisions.

References

Boldt, R. R., Comparison of a Bayesian and a least squares method of educational prediction (ETS RB 75-15) Princeton, NJ: Educational Testing Service, 1975.

Burns, R. L., Graduate admissions and fellowship selection policies and procedures, Part 1 and Part 2 (Supplementary Tables). Princeton, NJ: The Graduate Records Examination Board and Educational Testing Service, 1970.

Carlson, A. B., Evans, F. R., & Kuykendall, N. J., The feasibility of common criterion validity studies of the GRE (ETS RM 73-16). Princeton, NJ: Educational Testing Service, 1973.

Creager, J. A., A study of graduate fellowship applicants in terms of Ph.D. attainment (Technical Report No. 18). Washington, DC: Office of Scientific Personnel, National Academy of Sciences--National Research Council, 1961.

Creager, J. A., Predicting doctorate attainment with GRE and other variables (Technical Report No. 25). Washington, DC: Office of Scientific Personnel, National Academy of Sciences--National Research Council, 1965.

GREB Newsletter (No. 20), Spring 1975.

Harvey, P. R., & Lannholm, G. V., The performance of foreign graduate students on the Graduate Record Examinations Aptitude Test (GRE Special Report 61-1). Princeton, NJ: Educational Testing Service, 1961.

Harvey, P. R., & Pitcher, B., The relationship of Graduate Record Examinations Aptitude Test scores and graduate school performance of foreign students in four American graduate schools (GRE Special Report 63-1). Princeton, NJ: Educational Testing Service, 1963.

Lannholm, G. V., Abstracts of selected studies of the relationship between scores on the Graduate Record Examinations and graduate school performance (GRE Special Report 60-3). Princeton, NJ: Educational Testing Service, 1960.

Lannholm, G. V., Review of studies employing GRE scores in predicting success in graduate study, 1952-1967 (GRE Special Report 68-1). Princeton, NJ: Educational Testing Service, 1968.

Lannholm, G. V., Summaries of GRE validity studies, 1966-1970 (GRE Special Report 72-1). Princeton, NJ: Educational Testing Service, 1972.

References (cont'd.)

Lannholm, G. V., & Schrader, W. B., Predicting graduate school success--An evaluation of the effectiveness of the Graduate Record Examinations. Princeton, NJ: Educational Testing Service, 1951.

Lannholm, G. V., Marco, G. L., & Schrader, W. B., Cooperative studies of predicting graduate school success (GRE Special Report 68-3). Princeton, NJ: Educational Testing Service, 1968.

Linn, R. L., Fair test use in selection. Review of Educational Research, 1973, 43, 139-161.

Linn, R. L., Test bias and the prediction of grades in law school, Journal of Legal Education, 1975, 27, 293-323.

National Academy of Sciences--National Research Council. Doctorate recipients from United States universities, 1958-1966 (NAS Publication 1489). Washington, DC: NAS--NRC, 1967.

Reilly, R. R., Critical incidents of graduate student performance (GREB Technical Memorandum No. 1). Princeton, NJ: Educational Testing Service, 1971.

Reilly, R. R., Factors in graduate student performance (ETS RB 74-2). Princeton, NJ: Educational Testing Service, 1974.

Rock, D. A., & Harmon, L. R., The identification of population moderators and their effect on the prediction of doctorate attainment. Unpublished manuscript, Educational Testing Service, 1972. (a)

Rock, D. A., & Harmon, L. R., The prediction of doctorate attainment in psychology, mathematics, and chemistry (GREB Preliminary Report). Princeton, NJ: Educational Testing Service, 1972. (b)

Rubin, D. B., Using empirical Bayes techniques in the law school validity studies. Report #LSAC 78-1a. Princeton, NJ: Law School Admission Council, 1978.

Schrader, W. B., The predictive validity of College Board admissions tests. In Angoff, W. H. (ed.), The College Board Admissions Testing Program. Princeton, NJ: College Entrance Examination Board, 1971, 117-146.

Schrader, W. B., Summary of law school validity studies, 1948-1975, Report #LSAC 76-8. In Law School Admission Council, Reports of LSAC Sponsored Research: Volume III, 1975-77. Princeton, NJ: Law School Admissions Council, 1977, 519-550.

References (cont'd.)

Schrader, W. B., & Pitcher, B., Discriminant function analysis in the prediction of law school performance (ETS Statistical Report 72-40). Princeton, NJ: Educational Testing Service, 1972.

Sharon, A. T., Test of English as a foreign language as a moderator of Graduate Record Examinations scores in the prediction of foreign students' grades in graduate school (ETS RB 71-50). Princeton, NJ: Educational Testing Service, 1963.

Tucker, A., Gottlieb, D., & Pease, J., Attrition of graduate students at the Ph.D. level in the traditional arts and sciences (Publication No. 8). East Lansing, MI: Michigan State University, Office of Research and Development and the Graduate School, 1964.

Wainer, H., Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Wainer, H., On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

Warren, J. R., Prediction of college achievement among Mexican-American students in California (ETS RDR 76-77 & RB 76-22). Princeton, NJ: Educational Testing Service, 1978.

Willingham, W. W., Predicting success in graduate education. Science, 1974, 183, 273-278.

Willingham, W. W., Predicting success in graduate education. Papers presented at the Graduate Record Examinations Board Research Seminar, 12th Annual Meeting of the Council of Graduate Schools. Princeton, NJ: Educational Testing Service, 1973.

Wilson, K. M., Predicting the long-term performance in college of minority and nonminority students: A comparative analysis in two collegiate settings (ETS RDR 77-78, No. 3 & RB 78-6). Princeton, NJ: Educational Testing Service, 1978.

Wilson, K. M., Of time and the doctorate--Report of an inquiry into the duration of doctoral study, Research Monograph No. 9. Atlanta, GA: The Southern Regional Education Board, 1965.

Wilson, K. M., The utility of a standard composite for forecasting academic performance in several liberal arts colleges. Research in Higher Education, 1976, 5, 193-213.

Wilson, K. M., The contribution of measures of aptitude (SAT) and achievement (CEEB Achievement Test Average), respectively, in forecasting college grades in several liberal arts colleges (ETS RB 74-36). Princeton, NJ: Educational Testing Service, 1974.

Appendix A

GRE Program Related Efforts to Promote Validation Research: Review and Appraisal

Appendix A (Page 1 of 14 pages)

GRE PROGRAM-RELATED EFFORTS TO PROMOTE VALIDATION RESEARCH: REVIEW AND APPRAISAL

During the past quarter of a century, those concerned with the development and use of the Graduate Record Examinations (GRE) have periodically called attention to the need for reliable knowledge bearing on the validity of the GRE tests in the selection of individuals for admission to graduate programs. The need to improve the validity study process in graduate schools has been endorsed and supported.

A number of projects undertaken by ETS (with the encouragement of either the Committee on Testing of the Association of Graduate Schools [AGS] or, more recently, the Graduate Record Examinations Board [GREB]) have been designed (a) to improve the quality of information available regarding the validity of the GRE tests, (b) to advance understanding of the validity study process, and/or (c) to focus attention on special problems or promising developments. The projects undertaken to date have been of three types:

a) those concerned with periodically collecting and disseminating information and insights gained from institutional, departmental, and other validity studies,

b) those concerned with designing and conducting validity studies with the cooperation of individual institutions or departments, and

c) those concerned with the study of particular applied, methodological, or conceptual aspects of the validation process or with particular problems and issues.

A review of these three types of effort provides useful perspective for developing a strategy for improving the validity study process (which has not become an established aspect of institutional operations at the graduate level).

Collecting and Disseminating Information

In the first summary of information on institutional or departmental validity studies, Lannholm and Schrader (1951)* described major studies of the prediction of graduate school success by the

---

*See consolidated references following main body of the report.

GRE tests that were reported to the GRE Office during the period 1937 through 1951. The investigators concluded that carefully constructed tests of achievement in major subject matter fields may be used effectively in the admission and guidance of graduate students and that the Advanced Tests might be given precedence over the then available Profile Tests for purposes of predicting success.

The next summary of validity studies was not forthcoming until November 1960, when Lannholm (1960) summarized results of a limited number of validity studies in a report designed (a) "to illustrate different approaches to a study of [the relationship between scores on the GRE and success in graduate study] and (b) to stimulate other graduate schools to design and carry out studies of their own." Authors of unpublished studies were invited to send copies of their reports to ETS.

A more comprehensive summarization appeared eight years later. Lannholm (1968) reported on 36 studies conducted over a 15 year period--i.e., between 1952 and 1967--thirteen of which were in the field of Education. In examining the various approaches taken, Lannholm noted that most of them involved the analysis of data separately for samples by discipline or department but that a few studies pooled data for samples from several disciplines or departments. The only study included in this summary which involved the pooling of data from more than one graduate school was a study by Creager (1965) of the relationship between GRE scores and several related doctorate-attainment criteria in a national sample of applicants for NSF fellowships.

In the 1968 report, attention was focused squarely on the "persistence of certain problems" in connection with designing and conducting validity studies in graduate schools. Lannholm cited as the principal problems those related to (1) the small size of samples, (2) the lack of a single satisfactory index of the effectiveness of predictors, and (3) the limitations of measures of graduate school success. The perceived need to conduct validity studies by department contributed to restriction of sample size; difficulties involved in interpreting correlation coefficients in highly restricted ranges of talent were held to militate against the routine use and interpretation of familiar correlational procedures for assessing predictor effectiveness; and the limitations of grade point averages, frequently employed as criterion measures, were cited. Lannholm characterized as "...both surprising and disappointing..." the failure of most investigations to include a measure of undergraduate performance (e.g., undergraduate grade point average) as a predictor.

In the most recent summary report, Lannholm (1972) presented the results of 14 studies, received by the GRE Office after the 1968 report had been prepared, for the period 1966-1970.

Again, emphasis was placed on continued problems of sample size. Efforts to enlarge sample size by pooling data for two or more departments were noted. However, Lannholm noted that even though "...larger numbers result from pooling data from different departments, the effect upon the prediction coefficients is difficult to determine, especially when the abilities required and the performance standards vary from one department to another." An undergraduate grade point average was used in conjunction with test scores in ten of the 14 studies; it was suggested that adjustments for quality of the undergraduate institution seemed promising. The need for further work on the development of satisfactory criteria of success in graduate study was stressed.

No comparable summarization of the results of departmental validity studies has been reported since 1972. However, interest in this line of endeavor continues; the Spring 1975 issue of the GRE Board Newsletter included a request that graduate schools forward reports of validity studies carried out within the past five years.

The collection and dissemination of data on validity provided by local, institutional and departmental studies clearly constitutes a necessary element in a comprehensive plan for improving the validity study process. However, the fact that this approach is not sufficient has been recognized; other approaches that have been supported are considered in the following sections.

Cooperative Validity Studies

In 1962, recognizing the limitations of many institutional-departmental validity studies, the Committee on Tests of the AGS recommended that ETS undertake validity studies in cooperation with several graduate schools. In 1963, 32 departments in 15 different universities were invited to participate in studies designed to (a) evaluate the effectiveness of scores on the GRE and other factors in predicting success in graduate study in certain departments in selected graduate schools, and (b) to provide suggestions to other departments and schools that might wish to study the effectiveness of their own selection procedures (Lannholm, Marco, and Schrader, 1968).

The "cooperative validity studies" approach represented an important advance. Local studies were seldom strictly comparable in design and methodology and they typically employed different kinds of criterion measures and samples. The cooperative validity studies model, on the other hand, employed a standard methodology and design. Analyses were centrally planned and conducted, while the departments cooperated in supplying the necessary data.

Data were received on samples of students from 21 departments, spanning six fields at ten universities. The studies were conducted on a departmental basis; the design did not call for pooling data from different departments within the same discipline. Comparable sets of predictor-criterion data were generated for many of the departments, however.

These studies used a complex criterion variable defined in such a way as to reflect (a) the "progress" of individuals through various aspects of the "general excellence" of students. "Progress during the study period" (and status at the time of the cut-off date involved) was reported for each student as follows:

a. earned Ph.D.
b. passed all examinations, still enrolled
c. has not passed all examinations, still enrolled
d. withdrew at the request of the university (dismissed)
e. voluntarily withdrew after more than a year of study
f. voluntarily withdrew after less than a year of study

Departmental ratings were employed along with the foregoing categories to define "successful" and "unsuccessful" groups as follows:

> "successful" students were those who had received the degree or who, if still enrolled, had highest ratings of "outstanding" or "superior."
>
> "unsuccessful" students were those who had not received the degree, who were still enrolled with "average or lower ratings, or who had withdrawn regardless of circumstances.

Only two of the samples studied included more than 85 individuals; the small samples no doubt contributed to the variation in results from one group to another. It was evident from the study, not only that the validity of the GRE and undergraduate predictors varied considerably,* but also that there was marked variability among the departments with respect to the distribution of students according to status at the end of the study period (see Table A-1). In some

---

*In retrospect, several factors may have contributed to variation in results from one sample to another and have had an attenuating effect on the validity coefficients obtained. Among these factors are the classification as "unsuccessful" of students who withdrew voluntarily and the inability, due to small sample size, to analyze data separately by sex. The number of males and females involved was not reported. However, the criterion involved had as one of its elements "degree attainment within a specified time period." Women have tended to take longer to complete degree requirements due to a number of non-ability sex-role-linked factors. Analyses by sex are important.

Insert Table A-1 about here

departments with respect to the distribution of students according to status at the end of the study period (see Table 1). In some departments, for example, almost two-thirds of the sample had attained the Ph.D.; in others from 40 to 94 percent had not passed all examinations; in still others, half or more of the students had either withdrawn voluntarily or been dismissed.*

Lannholm, Marco, and Schrader (1968, p. 84) concluded that "on the whole, the results of these studies make it clear that prediction of success in graduate work is exceedingly difficult... and that in view of the critical importance of graduate study and the importance of effective prediction both to the student and to the graduate school, the results emphasize the urgency of seeking ways to improve prediction."

As previously noted, the 1968 Cooperative Validity Studies (launched in fall, 1963) embodied the principle of applying a standard study design to data for each of several departments. ETS investigators were responsible for study design, while the departments were asked to cooperate by supplying specified data for the study; and the investigators were successful in obtaining the cooperation of 10 of the 15 schools and 21 of the 32 departments originally invited to supply data for the study. Good cooperation was obtained from faculty members in participating departments in supplying ratings of students.

A second graduate-level project calling for the cooperation of departments in supplying data for a centrally designed validity-related study, was initiated with GREB sponsorship in 1970 (Boldt, 1975). This study was designed to examine the utility of "special new statistical techniques" (Bayesian analysis) for weighting a common set of predictors in several prediction contexts where small sample size tends to be a problem. Eighty-one departments of psychology and 54 departments of economics were approached. Despite the fact that this study was endorsed by the GREB Chairman through a covering letter to the graduate deans, the research could not be carried out due to the fact that most of the invited departments did not provide the required data (Boldt, 1975, pp. 12-14). The few which did supply data had limited samples.

---

*See Wilson (1965) for evidence of marked variability among departments in rate of progress of students in completing doctoral programs.

Table A-1

Departmental Differences in Patterns and Rates of Progress in Ph.D. Study

Field/ Department		Status as of study cut-off date* (in percent)**						
		Still Enrolled				Dismissed	Voluntarily Withdrawn	
		N	Ph.D. completed %	Exams passed %	Exams not passed %	%	Later %	Earlier %
Chemistry	A	(116)	37.9	19.8	0.0	8.6	7.8	25.9
	B	(20)	65.0	5.0	0.0	0.0	10.0	20.0
Total		(136)	41.9	17.6	0.0	7.4	8.1	25.0
English	C	(98)	25.5	41.9	8.2	6.1	2.0	16.3
	D	(81)	13.6	2.5	46.9	13.6	12.3	11.1
	E	(54)	31.5	11.1	7.4	0.0	29.6	20.4
	F	(32)	28.1	6.2	40.6	9.4	9.4	5.9
Total		(265)	23.4	19.2	23.8	7.5	11.7	14.3
History	G	(66)	12.1	1.5	33.3	19.7	13.8	19.7
	H	(40)	7.5	55.0	22.5	5.0	2.5	7.5
	I	(28)	28.5	35.7	0.0	3.6	10.7	21.4
Total		(134)	14.2	24.6	23.1	11.9	9.7	16.4
Philosophy	J	(42)	38.1	33.3	0.0	0.0	11.9	16.7
Physics	K	(39)	43.6	23.1	0.0	0.0	17.9	15.4
	L	(38)	0.0	2.6	94.7	0.0	2.6	0.0
	M	(32)	65.6	0.0	34.4	0.0	0.0	0.0
Total		(109)	34.9	9.2	43.1	0.0	7.3	5.5
Psychology	N	(49)	16.3	24.5	36.7	6.1	16.3	0.0
	O	(47)	42.6	4.3	19.1	4.3	2.1	27.7
	P	(44)	20.4	6.8	13.6	34.1	18.2	6.8
	Q	(38)	39.5	5.3	7.9	0.0	31.6	15.8
	R	(36)	63.9	16.7	19.4	0.0	0.0	0.0
	S	(26)	19.5	46.2	19.2	3.8	7.7	3.8
	T	(26)	50.0	42.3	3.8	0.0	3.8	0.0
Total		(266)	35.0	18.0	18.4	7.9	12.0	8.6
All fields/ departm[illegible]		(952)	29.9	18.9	20.0	7.0	10.5	13.7

*Compiled from Lannholm, Marco, and Schrader (1968), based on the October, 1963 status of students "...first enrolled between the fall of 1957 and June 1960..."

**Row totals should equal 100 percent within limits of rounding.

The methodological and exploratory nature of this particular study may have proved to be a deterrent to cooperation. Other considerations as well may have contributed to the failure of institutions to provide data, including, for example, the nature and availability of the data requested. The clerical burden involved and issues relating to the confidentiality of student data were cited frequently. Other reasons included the lack of availability of data on Ph.D.'s and the lack of availability of GRE scores due to the fact that the score-requirement policy was not actually enforced.

In any event, this particular "cooperative study" project failed to elicit the required cooperation of graduate departments in two fields of study. However, the methodological aspects of the investigation were completed successfully with data available in College Board's Validity Study Service files at ETS for a number of freshman samples. And, as will be considered later, the concepts underlying the design of the study have important implications for development of the validity study process in graduate-study contexts.

It is evident that the two projects reviewed above were only partially successful in generating validity study data. An expanded cooperative validity-study model calling for the participation of individual or defined groups, of departments in centrally coordinated and facilitated validity studies might include provision for institutional-developmental involvement in planning and designing, as well as in providing data for the studies. In any event cooperation and collaboration constitute necessary elements in any overall plan for facilitating the development of validity studies in graduate school settings.

Studies of Special Problems and Promising Developments

As previously noted, the project by Boldt (1975) did not elicit the cooperation of graduate departments. However, in its design Boldt introduced a number of ideas that have important implications for the validity-study process. In essence, Boldt focused attention on the need for approaches to the study of validity which assume that there are important elements of similarity in the tasks required of individuals in each of several different prediction contexts--e.g., several different graduate departments of chemistry--even though these tasks may be conducted at different levels, and with differing emphases.

Boldt (1975, pp. 1-2) offers the following relevant observations:

> In some population segments, such as minority groups, graduate students, and possibly various occupational groups, one often cannot find enough people at a single place where an acceptable criterion exists to conduct a statistical study of

> the predictive validity of selection instruments, or at least a study in whose results one can have confidence. It is more common to find small groups from the population of interest interspersed through a variety of locations, performing tasks that seem reasonably similar. Evaluation of the performances is made with reference to the group at a location but without reference to the performances outside that group... Thus, the groups may differ from each other in terms of average performance or in the variation in performance, but these differences may not be inferrable from the corresponding statistics calculated using quantitative evaluations of performances made at each location.
>
> Where several schools are involved, one would want to incorporate the notion that they are more or less similar. One would certainly _not_ want to proceed under the assumption that all schools are uniquely different, conceivably, and that no prior information [of value for facilitating the assignment of weights to predictors] is in existence.

The assumption that several graduate departments in a given field of study probably are engaged in a basically _similar_ enterprise suggests the possibility of improving validity studies through designs which call for the consideration of _common sets of predictor-criterion data on individuals in each of several "similar" departments_. This would result in a substantial enlargement of the data base available for analysis and enhance the generalizability of fundings.

GRE scores constitute a set of predictors which may be thought as common from one prediction context to another. However, the problem of establishing the "credibility" of a criterion variable (or variables) with comparable meaning across several graduate departments is not so readily solved. The "criterion problem" has been the focus of two recent GREB-supported projects, namely, Reilley's (1971, 1974) critical-incidents study of graduate-student behaviors aimed at "defining empirically a set of criterion dimensions upon which graduate faculty base judgments of student performance," and an exploration by Carlson, Evans, and Kuykendall (1973) of the feasibility of developing validity studies of the GRE, based on a "common criterion."

Reilley's investigation was designed to identify aspects of student behavior that might help define criterion dimensions which graduate faculty members could use in judging student performance. Procedures such as those developed by Reilley clearly should be useful in exploratory validation research. They provide a basis for taxonomic investigations as well--e.g., for clustering departments in terms of the types of student behavior deemed most important by the faculty.

Citing the concern of the GREB Research Committee over the "paucity of validity data for the GRE," and the interest of GRE Committees of Examiners in validity studies, Carlson et al. (1973) undertook an exploratory investigation of the feasibility of developing for one or more fields "a measurable criterion which would be generally acceptable to at least a large segment of that field," probably a common set of essay questions to be administered to students at the appropriate level in each of several different departments. Discussion of the possibilities of developing and using an essay-type measure with GRE Committees of Examiners in Philosophy, French, and Literature in English led to the conclusion that "the problems of such a study were insurmountable, and the procedure was rejected."

Reactions of the respective committees were varied. In Philosophy it was deemed feasible to obtain agreement on several essay questions, but the Committee doubted the adequacy of such questions as a criterion; they were unable to specify a task or set of tasks which they would find to be an acceptable criterion (though they felt that "rating scales offered real possibilities"). In French, interest was keen but efforts to implement the idea were not successful--many of the department chairmen indicated concern over the operational problems posed by introducing a special examination. For the Literature in English Committee, "essay questions" were not acceptable as a criterion for graduate student performance in their field. They did express interest in "attainment of tenure in a 'good' department" as a criterion (implicitly, "quality level" of the institution in which graduates were finally "placed"); problems involved in implementing this idea were explored briefly but it was not pursued further.

Although the standard-set-of-essay-questions approach to developing a common criterion was considered to be inappropriate, there was considerable interest in the possibilities involved in using rating procedures. Based on a survey of selected departments in five fields, some 43 percent of responding departments reported regular use of some form of rating, typically at master's or Ph.D. examination times.

The investigators concluded from the survey results that a sufficient number of departments were employing rating procedures to warrant some preliminary studies based on existing rating data but cautioned that it would probably be desirable to develop a uniform set of rating procedures before using ratings as criterion measures.

These explorations of the feasibility of using a common criterion yielded a negative conclusion only with regard to the feasibility of employing one particular form of "common criterion"--a common set of essay questions. The negative conclusion does not apply to the idea of a common criterion, per se, or the validity of an implicit assumption underlying the proposed use of such a criterion, namely, that the tasks required of students in different departments within a given field are sufficiently similar to warrant the use of a common criterion measure (or, that it is possible to identify a subgroup of departments which are by some acceptable means judged to be sufficiently similar in regard to demands made on students to warrant use of such a criterion measure).

An extension of GREB-supported research reflecting concern over the criterion problem is represented in the study by Rock (1972), with the collaboration of Lindsey Harmon, which used data obtained from the NSF Fellowship applicant records and the NRC Office of Scientific Personnel Doctorate Records File (DRF). The study was designed to evaluate the validity of GRE Aptitude and Advanced Tests as predictors of whether or not a candidate (in psychology, mathematics, or physics) attained the doctorate within a given period of time, extending and elaborating a line of inquiry associated with Creager (1961, 1965).

The GREB-sponsored study, like Creager's earlier studies, examined the validity of GRE tests vs. Ph.D. attainment in a sample undifferentiated with respect to institutional affiliation, but it also sought to determine whether there were particular subgroups within the fields under consideration for which the GRE test might have varying degrees of predictive validity. Examples of such subgroups are "quality level" of the graduate school, age at the beginning of study and sex. And the study provided, incidentally, relevant information bearing on the potential value of attainment vs. nonattainment of the Ph.D. "within a reasonable time" as an administratively practical "common" criterion (or component in such a criterion) in a validation model having both within-department and across-department components.

A rationale for use of the relatively crude Ph.D. attainment vs. nonattainment criterion was offered by Rock and Harmon:

> The most desirable criterion, of course, would be some measure of achievement as a scientist. Aside from the logical difficulties in arriving at any sort of agreement as to what is a relevant measure of scientific achievement, we are faced with the operational problem of time lapse which must occur before such data can be collected.
>
> An alternative criterion of a more intermediate nature is whether or not one has attained his or her doctorate within a reasonable period of time. Attainment of the doctorate is appealing on logical grounds since...it is one test of the effectiveness of the overall selection process, i.e., the decision to admit a student to graduate education or to admit him to candidacy for a higher degree implies an expectation that his formal graduate education will be completed. The attainment of the [doctorate degree] is the primary indicator that such an expectation has been fulfilled...One criticism... is that doctorate attainment lacks sensitivity in the sense that it cannot take into account the various qualitative levels of performance among individuals attaining the Ph.D.

> Although the latter criticism may well be valid... [if this criterion] is sufficiently lacking in sensitivity, this in turn will be reflected in the relative level of its predictability (pp. 1-2).*

The Ph.D. attainment criterion thus ignores "levels" within that classification, and it ambiguously assigns all other individuals to a nonattainment category. Some of these individuals will later become "attainers." Accordingly, the criterion should be perceived as reflecting differences among individuals in rates and patterns of progress in completing programs of Ph.D. preparation.

GRE scores have been found to be related, consistently if modestly, to criteria which reflect "rate of progress" toward the degree. Differences in "ability" may partially account for observed differences in average degree-attainment times by institutional attendance and degree pattern (see Table A-2). Differences in degree-attainment rates by "quality level" of institutions may also be ability-related (see Table A-3), and differences among several departments within a given field with respect to average rates of student progress may be accounted for partially by differences in "quality of student input." Thus, rate of student progress in completing degree requirements appears to have considerable promise as one component in a "common criterion" variable reflecting the progress of individuals in completing requirements for graduate degrees, especially the Ph.D.**

---

*Reliance on "raw" attainment vs. nonattainment criteria is understandable in studies involving samples not identified with particular institutions/departments and an available data base which does not permit the development of more refined criteria. In study designs which involve analyses both within- and across-institutions/departments (field constant), more refined criterion groupings may be developed so as to reflect in some appropriate combination, for example, "degree attainment vs. nonattainment," faculty ratings of the overall excellence of the work done by degree attainers, rate of progress in gaining admission to candidacy, failure to qualify, dismissal on grounds of inadequate academic performance, etc.

**In carefully controlled "rate of progress" designs, all individuals involved should be at the same stage of preparation at the beginning of a study period in order to have the same amount of time in which to attain the degree or to reach any specific level of preparation (e.g., completion of course requirements, admission to candidacy through qualifying examination, etc.). Studies using Ph.D. attainment vs. "nonattainment" typically have not adequately controlled the "equal time" variable. For example, in order to enlarge the sample, students enrolled during, say, a given three-year period, are included, but a uniform cut-off date typically is employed.

Table A-2

Basic Institutional Attendance and Degree Patterns for Ph.D. Recipients and Associated Measures of Degree Attainment Time*

Institutional and degree pattern	%	Index of time taken	
		Registered time mean years	AB to Ph.D mean years
Direct Ph.D.	[20.0]	[4.8]	[5.6]
AB = Ph.D.	(3.8)	4.7	5.4
AB ≠ Ph.D.	(16.2)	4.8	5.6
Master's-Ph.D. same school	[41.4]	[5.3]	[7.4]
AB = MA = Ph.D.	(12.8)	5.4	7.3
AB ≠ MA = Ph.D.	(28.6)	5.3	7.5
Master's-Ph.D. different school	[38.6]	[6.1]	[11.0]
AB = MA ≠ Ph.D.	(15.6)	5.9	9.8
AB ≠ MA ≠ Ph.D.	(23.0)	6.3	11.8

---

*Data from NAS (1967) for U. S. doctorates (excluding foreign degree recipients) for 1966. The variations in mean attainment times shown here for all degree recipients tend to hold for essentially all broad fields. As suggested in the NAS publication (1967, p. 77), this association "...may be caused by different student abilities in the different [institutional attendance and degree] patterns, but no data exist to verify this guess."

Table A-3

Selected Data on the Percentage of Individuals Attaining the Ph.D. within a Given Study Period by "Quality" of Institution/Department and Ability Levels

Data source and field	Percent attaining Ph.D. within study period by "quality level"			Total
	"Higher" %	"Middle" %	"Lower" %	
Rock and Harmon (1972)*				
Psychology	44	36	26	40
Mathematics	61	40	31	54
Chemistry	77	69	53	74
Tucker, Gottlieb, and Pease (1964)**				
Physical sciences	79	65	56	70
Biological sciences	80	67	49	71
Social sciences	68	53	44	59
Humanities	57	49	33	50
(All fields)	(70)	(57)	(46)	(62)
Creager (1965)***				
Biology (males)	62	35	30	37
Chemistry (males)	63	50	25	51
Mathematics (males)	37	13	0	24

*Study period: 1958-61 through June, 1968; "quality" based on Cartter and other data descriptive of graduate department in which enrolled

**Study period: 1950-53 through December 1962; "quality" or productivity defined as (1) top 15 universities in Keniston rankings, (2) 300 plus Ph.D.s awarded, 1936-56, but not top 15, and (3) less than 300 Ph.D.s awarded and not top 15. This was a study involving approximately 24,000 post-master students at 24 selected universities. The investigators concluded, in part, "...that to increase Ph.D. production and reduce attrition, graduate schools would embark on active programs of recruiting potential graduate students and be more selective in their admissions." (p. 293).

***Study period: 1954-57 through August 1964: quantitative ability levels (GRE Q) defined as follows—"higher" = stanines 8-9; "middle" = stanines 5-7; "lower" = stanines 1-4. (Table 6, p. 24, selected fields only, to illustrate trends).

Appendix B

B-1. Forms used in survey of deans of CGS-member graduate schools; covering letter form GRE Board chairman

B-2. "Survey of Graduate School Validity Study Activities and Interests: Summary of Findings"

A report of findings of the survey of deans

B-3. Basic one-year, two-cohort validity study model used for the Cooperative Studies

Statement regarding confidentiality of treatment of data

Study definitions, data collection procedures, etc.

B-4. Brief description of selected studies using a two-year, single-cohort study model

Study definitions, data collection procedures, etc.

Appendix B-1

Graduate Record Examinations Board
PRINCETON NEW JERSEY 08540 • AREA CODE 609 921-9000

IN AFFILIATION WITH
The Association of Graduate Schools
The Council of Graduate Schools

April 13, 1976

1975-1976

Sanford S. Elberg
University of California at Berkeley
Chairman

•

Richard Armitage
Ohio State University

May Brodbeck
University of Iowa

Beverly Cassara
Federal City College

David R. Deener
Tulane University

I. Wesley Elliott
Fisk University

Wytze Gorter
University of Hawaii at Manoa

Arthur F. Jackson
North Carolina Agricultural
and Technical State University

Lyle V. Jones
University of North Carolina
at Chapel Hill

Sterling M. McMurrin
University of Utah

J. Boyd Page
Council of Graduate Schools
(ex officio)

George J. Resnikoff
California State University
at Hayward

Lorene L. Rogers
University of Texas
at Austin

Ben Rothblatt
University of Chicago

Harry H. Sisler
University of Florida

Donald J. White
Boston College

W. Dexter Whitehead
University of Virginia
at Charlottesville

Maryann A. Lear
Secretary to the Board

Dear Colleague:

To help meet the need for current information regarding GRE validity-study activities and to facilitate the development of more systematic and regular procedures for assessing the predictive validity of the GRE, the Graduate Record Examinations Board has funded a three-year project designed to achieve these goals. The project is intended to encourage and facilitate GRE validity-study research in graduate school settings where a variety of complexities--organizational, conceptual, statistical, and logistical--have made it difficult for concerned deans and faculty members to design and conduct such studies in the past.

Briefly, graduate schools willing and able to provide necessary data may obtain assistance from Educational Testing Service in designing validity studies. ETS will also analyze the data and report findings; institutions participating in this cooperative effort will receive copies of the results of the research. Multi-institutional/departmental approaches to GRE validation research will be explored, e.g., studies involving the concurrent participation of departments from the same set of fields at each of several cooperating graduate schools using a standard design and comparable data. In reporting about GRE validity studies, the information supplied by graduate schools will not be identified with a particular institution and will be held confidential.

The information called for in the enclosed two-part questionnaire is critical for the planning and development of a cooperative effort. It is needed to identify institutions/departments that have conducted GRE validity studies in recent years and to identify those interested in exploring actively the possibility of participating in the cooperative effort. If you report an interest in participating, appropriate follow-up inquiries will be made; even if you are not interested in further validity work at this time, your completion of this questionnaire will be of great value.

Your assistance in completing the questionnaire and in sharing the results of any institutional/departmental validity studies that have been completed since 1970 will be greatly appreciated and will contribute substantially toward the goals set by the Board in funding this important project.

Sincerely yours,

Sanford S Elberg

Sanford S. Elberg
Chairman

Enclosure

cc: Maryann A. Lear

GRE No. __________(ETS use)

GRADUATE RECORD EXAMINATIONS BOARD

A Survey of GRE Validation Research Activities and Interests Part 1

Name of institution ______________________________ Location ____________________

Name of respondent ______________________________ Title ____________________

* * * * * * * * *

About the Survey

This survey is part of the Cooperative Validity Studies Project being conducted by ETS for the Graduate Record Examinations Board.

Part 1 of the survey calls for (a) general classificatory information (e.g., size, control, highest degree), (b) limited data or estimates regarding graduate-school wide application/enrollment status/GRE-score availability variables, (c) general information about the incidence and nature of GRE validity-study activity since 1970, and (d) an assessment of the general level of interest and/or concern regarding questions related to the validity of GRE scores for predicting student performance.

Part 2 of the survey calls for information that will help to identify departments or programs (a) for which GRE validation research may be relevant, (b) in which validity studies have been completed or are in progress, and (c) which, from the perspective of the graduate dean's office, would be interested in exploring further the possibility of participating in a cooperative study. This information is critical from the point of view of study planning. Appropriate follow-up inquiries will be made to assess both interest and readiness to participate in studies.

For your reference, a copy of a recent GRE Program "Cumulative Summary Statistics Report" prepared for your institution has been included in this mailing. This report indicates the total number of GRE Aptitude Test score reports forwarded in a recent year as well as the number of Advanced Test score reports in up to 19 fields.

General Instructions

1. Please complete both parts of the survey at your earliest convenience.
2. Use the business reply envelope provided for returning the completed survey and any available validity study reports or summaries to Educational Testing Service.
3. If you have questions about the survey, call collect as follows:

Kenneth M. Wilson 609-921-9000, Ext. 2391
Educational Testing Service R208
Princeton, NJ 08540

Information provided in the survey will not be identified with your institution by name. It will be included in summaries for groups of institutions and departments.

IN ANSWERING QUESTIONS V - IX, PLEASE NOTE THAT BEST ESTIMATES ARE REQUESTED.

A Survey of GRE Validation Research Activities and Interests: PART 1--General

I. Highest level of degree awarded (or graduate study offered).

_____1 Doctorate
_____2 Master's
_____3 Beyond Master's, less than Doctorate

II. Institutional control or affiliation.

_____1 Public
_____2 Private, nonsectarian
_____3 Private, sectarian

III. Number of degrees awarded, 1974-75 academic year, including summer 1975.

Master's	Doctorate
_____1 None	_____1 None
_____2 1-149	_____2 1-149
_____3 150-299	_____3 150-299
_____4 300-499	_____4 300-499
_____5 500-999	_____5 500-999
_____6 1,000 +	_____6 1,000 +

IV. Does your institution have a general uniform admissions policy that applies to all graduate departments, i.e., not necessarily the same standard but a common policy?

_____1 Yes
_____2 No (Skip to Question V)

If "Yes," please check the statement below that best describes the general admissions policy of your institution.

_____1 Essentially "open door," i.e., all applicants who meet certain minimal requirements (such as holding a bachelor's degree) are admitted to pursue a graduate degree.

_____2 Essentially "open door" insofar as taking graduate courses is concerned but admission for degree purposes is a selective process.

_____3 Candidates meeting certain standards (e.g., specified undergraduate average and/or GRE-score minimum) may be admitted. Others may be admitted on an exception basis, even if below minimums.

_____4 Admission is on a competitive, comparative basis with other applicants seeking admission to a particular program for a given time period.

_____5 Other ______________________________
____
____

V. During the admissions year, 1974-1975, involving applicants for Fall 1975, how many applications for admission to the graduate school were received, how many applicants were accepted for admission, and how many accepted applicants actually enrolled? Consider degree-credit applicants only.

__________1 Total applicants, for Fall 1975
__________2 Number applicants accepted
__________3 Number accepted applicants enrolling

VI. Should an individual interested in applying for admission to your graduate school to pursue a degree program, submit GRE Aptitude and/or Advanced Test scores in connection with the application? Please select the one answer below that best reflects institutional/departmental practice (requirements, expectations, and the like) with regard to Aptitude and Advanced Test scores, respectively.

GRE		
Aptitude	Advanced	
_________1	_________1	Yes, scores should be submitted
_________2	_________2	No, scores need not be submitted
_________3	_________3	Answer depends on applicant's intended department/field/degree, undergraduate record, etc.

NOTE. Questions VII, VIII, and IX call for best estimates only for certain graduate school-wide statistics for variables that affect validity-study planning. The class intervals provided reflect relatively large tolerances for these estimates. If you have more precise data than called for by the categories provided, please check the broad category that is appropriate, and then enter the more precise figure in the space provided.

VII. From the perspective of the graduate dean's office, what is your best estimate of the proportion of applicants, graduate school-wide, for degree-credit enrollment in Fall 1975, submitting GRE Aptitude Test scores in connection with their application for admission?

_______1 Essentially all (or __________%)
_______2 90 percent plus
_______3 75-89 percent
_______4 50-74 percent
_______5 25-49 percent
_______6 10-24 percent
_______7 Less than 10 percent

VIII. Of all individuals enrolling for the first time in your graduate school in Fall 1975, what is your best estimate of the proportion classifiable as "first-time-enrolled graduate students"—i.e., with no previous graduate study at any institution? [Degree-credit only, full- or part-time]

_______1 Essentially all (or __________%)

_______2 90 percent plus

_______3 75-89 percent

_______4 50-74 percent

_______5 25-49 percent

_______6 Less than 25 percent

IX. Of all first-time-enrolled graduate students entering in a given year, what is your best estimate of the proportion likely to continue their enrollment into the second year, on either a full- or a part-time basis?

_______1 Essentially all (or __________%)

_______2 90 percent plus

_______3 75-89 percent

_______4 50-74 percent

_______5 25-49 percent

_______6 Less than 25 percent

> NOTE: Questions X, XI, XII, and XIII call for general information about GRE validity-study activities and interest. Please answer these general questions and then proceed to Part 2 of the survey which calls for more detailed information.

X. Have any institutional/departmental studies, designed in part at least to examine the relationship of GRE Aptitude and/or Advanced Test scores to any measure of student "success" in a degree program in any field(s), been completed at your institution since 1970? Are any such studies now in progress? Please enter one check (✓) in each column.

Validity Studies Completed?	In Progress?	
______1	______1	No (If "No" to both skip to XIII)
______2	______2	Yes, involving GRE Aptitude only
______3	______3	Yes, involving one or more Advanced Tests only
______4	______4	Yes, involving both Aptitude and Advanced Tests

XI. Have any of the studies completed or in progress been concerned directly or indirectly with the validity of GRE scores for predicting graduate school success among individuals in any of the following subgroups? Please answer for each subgroup by circling "yes" or "no."

Subgroup	Validity study involving GRE Aptitude?	Validity study involving GRE Advanced?
Women	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Black students	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Mexican-American students	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Puerto-Rican students .	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Other disadvantaged groups	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Older students, reentering the educational system	Yes$_1$ No$_2$	Yes$_1$ No$_2$
Part-time students . .	Yes$_1$ No$_2$	Yes$_1$ No$_2$

XII. Considering the validity studies that have been initiated and/or completed since 1970, at whose initiative were they undertaken? Indicate the individual, office, etc., primarily responsible for setting the studies in motion. If a single option will not suffice, check each applicable option.

_____1 The graduate dean and/or personnel associated with the dean's office

_____2 An office of institutional research

_____3 A central admissions office

_____4 A departmental chairman

_____5 A departmental committee

_____6 A standing committee of the graduate school

_____7 A student committee concerned with graduate school policies

_____8 A graduate student (thesis or dissertation)

_____9 An individual faculty member

_____10 An external agency ____________________

_____11 Other ______________________________

XIII. What is the current level of interest/concern in the graduate school over questions regarding the validity of GRE tests for predicting student performance in graduate study? Please provide your assessment of the general level of interest/concern from the point of view of (a) the dean's office, (b) the graduate faculty generally, and (c) student/applicant groups.

Office or Group	Level of interest/concern		
	Low	Medium	High
Graduate dean's office	1	2	3
Graduate faculty generally	1	2	3
Student/applicant groups	1	2	3

Please elaborate briefly below, indicating the types of questions and issues that are involved, reasons for concern or lack of concern, etc.

PLEASE COMPLETE PART 2

GRE No. ________(ETS use)

Name of institution______________________________Location______________________________

Name of respondent______________________________Title______________/Telephone__________

* * *

Survey of GRE Validation Research Activities and Interest **Part 2**

INVENTORY OF ACTIVITIES AND INTEREST

XIV. Is there at least one department/degree program in which at least half of all entering students ordinarily have GRE Aptitude and/or Advanced Test scores?

_____1 Yes (Please complete Questions 1, 2, 3, and 4, inside)

_____2 No (Please complete Question 1, inside)

General Instructions

The inventory may be completed by checking or entering a code for each of several designated fields in such a way as to indicate (a) the availability of a master's- and/or doctoral-level program, (b) GRE-score availability level, following the pattern suggested by XIV, above, and (c) whether validity studies have been completed or are in progress. You are also asked to identify programs or departments which, from the perspective of the graduate dean's office, may be potential participants in cooperative studies. Detailed instructions are provided inside. Please note the following general instructions:

1. In assessing GRE-score availability levels, best estimates only are sought.
2. Please provide copies of reports of completed validity studies whenever possible. If descriptions or summaries rather than copies of reports are deemed appropriate, please use the back of this inventory.
3. If you identify one or more departments or programs as possible participants in cooperative studies, appropriate follow-up inquiries will be made. No commitments are involved. If you are not in a position to specify particular departments or programs, but are interested in exploring further questions about participation in cooperative studies, indicate this by checking in the appropriate space inside.
4. If you have questions about the survey, call collect as follows:

   Kenneth M. Wilson 609-921,9000, Ext. 2391
   Educational Testing Service, R208
   Princeton, NJ 08540

5. When you have completed Part 2, please return both Part 1 and Part 2 (and copies of reports of studies, if available) to ETS in the business reply envelope provided for this purpose.

Information provided will not be identified with your institution by name. It will be used for study planning and in summaries for groups of institutions and departments.

SPECIAL NOTE: IF YOU HAVE CONDUCTED OR WISH TO CONDUCT GRE VALIDITY STUDIES IN FIELDS NOT LISTED INSIDE, PLEASE PROVIDE RELEVANT INFORMATION ON THE BACK OF THIS INVENTORY. AGAIN, PLEASE NOTE THAT ONLY BEST ESTIMATES ARE SOUGHT REGARDING GRE-SCORE AVAILABILITY.

General Inventory of GRE Validation Research Activities and Interests

LISTED BELOW ARE 19 FIELDS FOR WHICH A GRE ADVANCED TEST IS AVAILABLE. THESE FIELD DESIGNATIONS MAY BE THOUGHT OF AS REFERRING TO DEPARTMENTS/FIELDS/AREAS OF STUDY IN WHICH DEGREE PROGRAMS MAY BE OFFERED AS WELL AS TO PARTICULAR GRE ADVANCED TESTS. Answers to the questions included in this section of the survey will provide a comprehensive overview of the status of validity-study activities in the broad fields of graduate study listed and an inventory of institutional-departmental areas in which cooperative validity studies might be developed.

1. Is a degree program offered in the field? Check (✓) under M (Master's) in Column 1a and/or D (Doctoral) in Column 1b, as appropriate, to indicate at least one degree program at the designated level(s).

2. Is there at least one department/program in which GRE Aptitude and/or Advanced Test scores ordinarily are available for at least half the students entering each year? Check under "Aptitude & Advanced" in Column 2a to indicate one or more departments or degree programs in which half or more of enrolled students have scores on both Aptitude and Advanced Tests. Check under "Aptitude," Column 2b, to indicate one or more departments/programs in which half the students ordinarily have Aptitude scores but not Advanced Test scores. Check in Column 2c, under "Advanced" to indicate availability of Advanced but not Aptitude scores for a majority of students in one or more departments or programs.

   SPECIAL NOTE: ENCIRCLE A CHECK MARK IF THE LEVEL OF SCORE-AVAILABILITY FOR ANY DEPARTMENT/PROGRAM APPROACHES ESSENTIALLY COMPLETE COVERAGE (e.g., due to patterns of requirements).

3. For any program checked in 2a, 2b, and/or 2c, have institutional/departmental validity studies [to assess the degree of validity of GRE Test(s) for predicting student "success"] been completed since 1970? Are any validity studies in progress? USE "C" TO DENOTE A COMPLETED STUDY and/or "P" TO DENOTE A STUDY IN PROGRESS. Report in Column 3a those studies that involved both the Aptitude and an Advanced Test. In Column 3b, report studies involving the Aptitude Test only. Studies involving only an Advanced Test should be reported in Column 3c. Studies involving the GRE Aptitude Test in samples that are not homogeneous with regard to field/department (e.g., students from several social science departments) should be reported in spaces provided under Column 3b in the last three rows of the form, below.

   SPECIAL NOTE: FOR COMPLETED STUDIES PLEASE PROVIDE A COPY OF EACH REPORT OR A BRIEF SUMMARY OF STUDY DESIGN AND FINDINGS. FOR STUDIES IN PROGRESS PLEASE PROVIDE A BRIEF DESCRIPTION. Encircle "P" and/or "C" entries if you are including a validity-study report, summary, or description with your survey form.

4. A GRE validity study may be relevant for a department or program if "score availability" has been indicated in Columns 2a, 2b, and/or 2c." From the perspective of the graduate dean's office, in which of the relevant institutional/departmental areas is there currently active interest in the validation or further validation of GRE tests as predictors of student performance? Designation of an area as actively interested involves no commitment, of course, but should reflect the dean's judgment of institutional-departmental readiness to explore actively the possibility of participating in cooperative GRE validity studies GIVEN mutually acceptable study models and procedures. IN COLUMN 4, WRITE IN THE NAME(S) OF ALL ACTIVELY INTERESTED DEPARTMENTS/PROGRAMS. Indicate whether departmental/programmatic emphasis is on admission/selection/GRE validation for Master's study (M), Doctoral study (D), or both (M & D), by adding the appropriate letter(s) after the name of the department/field/program. If interested in possibility of a validity study, but unable to name specific areas, check space provided in 4a.

For detailed instructions, please refer to Questions 1, 2, 3, and 4, respectively, above.

Field/ department/ area	Degree program(s)?		GRE scores available, any dept./program?			Validity study since 1970?			Institutional/departmental areas interested in validation or further validation of GRE Aptitude and/or Advanced Tests? Enter name(s) of departments/ programs and M, D, or M & D as appropriate. [See also category 4a) at bottom of form].
	M	D	Apt. & Adv.	Apt.	Adv.	Apt. & Adv.	Apt.	Adv.	
	(1a)	(1b)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)
24 BIOLOGY									
27 CHEMISTRY									
31 ECONOMICS									
34 EDUCATION									
37 ENGINEERING									
44 FRENCH									
52 GERMAN									
91 SPANISH									
46 GEOGRAPHY									
47 GEOLOGY									
52 HISTORY									
57 LITERATURE English									
67 MATHEMATICS									
71 MUSIC									
74 PHILOSOPHY									
77 PHYSICS									
79 POLIT SCIENCE									
81 PSYCHOLOGY									
87 SOCIOLOGY									

COMBINED FIELDS: TO BE USED FOR REPORTING RECENT VALIDITY STUDIES IN 3b, GRE Aptitude	
11 Sample from two or more natural science fields	
12 Sample from two or more social science fields	
13 Sample from two or more humanities fields	

4a) [________] Check here if interested in exploring possibility of participating in a cooperative study, but not in a position to identify particular departments or programs at this time.

2.3

XV. TO WHOM SHOULD FOLLOW-UP INQUIRIES ABOUT VALIDITY STUDIES BE ADDRESSED? ____1 Respondent named on cover page

____2 ________________________

Name/Title/Telephone

2.4

Please use this page for descriptions of validity studies, for elaborating answers to questions, or for identifying validity study areas not covered by the inventory.

Appendix B-2

SURVEY OF GRADUATE SCHOOL VALIDITY STUDY ACTIVITIES AND INTERESTS: A SUMMARY OF FINDINGS

Information needed for project planning and development was obtained through a survey of graduate deans of institutions comprising the membership of the Council of Graduate Schools (CGS). The survey was designed primarily: (a) to identify prospective participants in cooperative validity studies, (b) to ascertain the types of questions or issues about GRE validity that were current oncampus, and; (c) to obtain information regarding current and recent levels of validity study activity, especially since 1970.

With a covering letter from the Chairman of the Graduate Record Examinations Board, the survey was mailed in April, 1976 to deans of 344 CGS member schools.* A total of 244 deans (or their representatives) responded. Some degree of interest in the possibility of participating in cooperative studies was indicated by 130 of the respondents. The role of the survey in identifying prospective participants in GRE validity studies is considered in a subsequent section. However, attention is directed first to information provided by the survey regarding the status of GRE validation research in CGS member schools.

Status of GRE Validation Research

The fact that a large number of schools (i.e., 130) indicated some degree of interest in the possibility of participating in cooperative GRE validity studies may be understood best when considered in relation to the extremely low incidence and uneven nature of local, institutional/departmental validity study activity reported by survey respondents. The survey included questions regarding (a) the extent of validity study activity since 1970, (b) studies that may have been completed for subgroups defined in terms of variables such as sex and/or ethnic group membership, and (c) the individuals or offices responsible for initiating and conducting the studies that had been made or were underway. In addition, it was requested that materials descriptive of completed or current studies be forwarded.

Judging from the responses to these questions, summarized in Table 1, and the nature of the descriptions and reports forwarded,

Insert Table 1 about here

---

*A copy of the covering letter and the survey forms used are included in Appendix B-1 (q.v.).

Table 1

Data on GRE Validity Study Activity Since 1970 in 244 CGS-Member Schools

---

Have any institutional/departmental studies, designed in part at least to examine the relationship of GRE Aptitude and/or Advanced Test scores to any measure of student "success" in a degree program in any field(s), been completed at your institution since 1970? Are any such studies now in progress? Please enter one check (·) in each column.

No. reporting

Validity Studies		
Completed?	In Progress?	
167	177	No
18	17	Yes, involving GRE Aptitude only
0	3	Yes, involving one or more Advanced Tests only
10	10	Yes, involving both Aptitude and Advanced Tests
(49)	(37)	No answer

Have any of the studies completed or in progress been concerned directly or indirectly with the validity of GRE scores for predicting graduate school success among individuals in any of the following subgroups? Please answer for each subgroup by circling "yes" or "no."

Subgroup	Validity study involving GRE Aptitude?		Validity study involving GRE Advanced?	
Women	Yes	9	Yes	4
Black students	Yes	7	Yes	1
Mexican-American students	Yes	3	Yes	1
Puerto-Rican students .	Yes	3	Yes	1
Other disadvantaged groups	Yes	2	Yes	1
Older students, reentering the educational system	Yes	4	Yes	4
Part-time students . .	Yes	4	Yes	2

Considering the validity studies that have been initiated and/or completed since 1970, at whose initiative were they undertaken? Indicate the individual, office, etc., primarily responsible for getting the studies in motion. If a single option will not suffice, check each applicable option.

11	The graduate dean and/or personnel associated with the dean's office
1	An office of institutional research
3	A central admissions office
13	A departmental chairman
1	A departmental committee
0	A standing committee of the graduate school
2	A student committee concerned with graduate school policies
3	A graduate student (thesis or dissertation)
	An individual faculty member
	An external agency ____________
	Other ____________
8	Dean/departmental committee
2	Dean/chairman
1	Dean/Office of institutional research

only a few graduate schools have conducted systematic studies of the predictive validity of GRE scores (and possibly of other preadmissions variables, such as undergraduate GPA) in any program during the period 1970-76.

Specifically, the survey asked, "Have any institutional/departmental studies, designed in part, at least, to examine the relationship of GRE Aptitude and/or advanced Test scores to any measure of student 'success' in a degree program in any field(s), been completed at your institution since 1970? Are any such studies now in progress?"

- Only 38 respondents indicated that studies involving either the GRE Aptitude only (18 schools) or both the Aptitude and Advanced Test (10) had been completed; studies in progress were reported by 30 schools, including some of the schools that reported completed studies--17 involved the Aptitude Test only, 3 an Advanced Test only, and 10 involved both the Aptitude and an Advanced Test.

The question which elicited this response framed a very "inclusive" definition of "validity study." It was intended to permit an affirmative response if any empirical examination of variation in student performance by GRE score-levels, or vice versa, had been undertaken, and documented sufficiently to warrant circulation intra-institutionally (e.g., as a memorandum, report, or tabular summary).

Studies not comprehensive. Both the low incidence of reported validity study activity and the uneven nature of the exhibits forwarded as descriptive of that activity reflect the essentially undeveloped state of the "validity study art" in graduate school settings.

- Only 10 survey respondents included materials descriptive of completed, current, or planned local studies of GRE predictive validity. Of the ten exhibits forwarded, only one involved both a systematic analysis of relationships among clearly defined criterion and predictor variables, and samples broadly representative of the respective functional divisions of the graduate school.

- The exhibits differed markedly in format, comprehensiveness of reporting, and classifiability as "validity studies." Materials forwarded as illustrative of local GRE validity study activity included, for example, two summaries of descriptive statistics on grades and GRE scores, by department, that did not consider relationships among the data elements described. Also included was a Xerox copy of a computer printout of a table of intercorrelations for one sample; a scatterplot of GRE scores and GPA and a brief memorandum commenting on an observed correlation in one sample in one field; excerpts from a graduate student's methods-project thesis report; two ad hoc studies in schools of Education; reprints of published study reports based on students entering during the 1960s.

Subgroup studies not available. In view of the occasional, ad hoc nature and limited scope of the "validity study process" generally, as inferrable from the foregoing analysis, it is entirely understandable that only a few respondents reported activity designed to sned light on more complex questions that arise regarding the comparative predictive validity of GRE scores (or other admissions variables) for groups defined in terms of sex, disadvantaged status, age and/or degree of continuity of graduate study, enrollment status (full-versus-part-time), etc.

- As indicated in Table 1, only a handful of schools reported that any of the studies undertaken since 1970 had been concerned directly or indirectly with questions regarding the predictive validity of GRE Aptitude scores for women (9 schools), black (7), Mexican-American (3), Puerto Rican (3), "other disadvantaged"(2), older (4), or part-time (4), students, respectively. Even fewer schools reported examination of the validity of GRE Advanced Tests for such subgroups. None of the exhibits forwarded involved analyses by subgroup.

Responsibility for Studies

It is reasonable to infer from the foregoing that a "validity study function" continues to be an undeveloped area in graduate schools generally--i.e., validity studies are not conducted regularly as part of a process having clearly perceived organizational, conceptual, and operational parameters. Other survey findings support this inference. For example, the few studies that have been undertaken reportedly were initiated by a variety of different individuals and offices (Table 1):

- The graduate dean and/or personnel associated with the dean's office were designated as primarily responsible for 22 of the studies completed or underway.

- 13 studies were initiated by departmental committees, 3 by departmental chairman, 3 by individual faculty members, and 2 by graduate students as projects associated with the programs of study.

- An office of institutional research or testing (evaluation research) and a central admissions office were cited one time each as involved in current or completed studies.

The graduate dean's office was the single most frequently cited initiator of GRE validity studies, but representatives of schools or departments within the graduate school were reportedly responsible in almost as many cases. The complex, decentralized nature of the graduate school clearly has militated against the development and implementation of a graduate-school wide approach to examination of the validity of admissions decisions based on GRE scores or other evidence of the qualifications of candidates for admission to graduate study. At the same time, it would appear that the graduate dean's office tends to have a higher degree of interest in questions bearing on the validity of GRE tests (and other data) for predicting student performance than representatives of the respective departments.

Current Questions and Issues Regarding GRE Predictive Validity: Deans' Assessments

The survey sought information regarding some of the specific questions and issues pertaining to GRE predictive validity that are currently of interest and concern to graduate schools as viewed from the perspective of the graduate dean. In addition, deans were asked to assess the general level of interest and concern regarding these questions (a) in the dean's office, (b) on the part of departmental faculty, and (c) in student/applicant groups.

As indicated in Table 2, only 33 deans reported a "low" level of interest while 94 reported a "high" level of interest in GRE validity-related questions; they perceived a somewhat lower level of interest in such questions among graduate faculty generally and in student/applicant groups.

Insert Table 2 about here

About 150 of the respondents provided some elaborative commentary in connection with their assessments of the general levels of interest and concern regarding GRE predictive validity. Many of the comments were relatively general in nature, referring to local patterns of GRE use rather than to validity-related concerns. For example:

Table 2

Question Regarding Dean's Perception of Level of Concern Over GRE Validity, and Associated Distribution of Responses

What is the current level of interest/concern in the graduate school over questions regarding the validity of GRE tests for predicting student performance in graduate study? Please provide your assessment of the general level of interest/concern from the point of view of (a) the dean's office, (b) the graduate faculty generally, and (c) student/applicant groups.

Not responding	Office or Group	Level of interest/concern		
		Low	Medium	High
13 (5.3%)	Graduate dean's office	33 (13.5%)	104 (42.6%)	94 (38.5%)
13 (5.3%)	Graduate faculty generally	60 (24.6%)	131 (53.7%)	40 (16.4%)
38 (15.6%)	Student/applicant groups	95 (38.9%)	78 (32.0%)	33 (13.5%)

> "Some departments place more emphasis on the Graduate Record Examinations than others--some have a cutoff score and will not accept applicants who do not meet this requirement."
>
> "Most scores are used here as a basis for admission since there is a heavy demand for places."
>
> "Although GRE scores are required from all degree program applicants, other factors are given equal weight in predicting chances for success in the Graduate School."

Reference to "opinion," "belief," or "conviction" regarding the "usefulness," "value," "validity," etc., of GRE scores was a frequently recurring element in other general comments emanating from graduate school settings in which no validity-study activity was reported.

> "Since the Graduate Faculty is unwilling to impose a university-wide requirement for the GRE, the interest in validity studies is limited. Departments who use it think it is valid; those who don't, think little about it..."
>
> "Many faculty members do not feel the GREs reflect the students' predictable performances."
>
> "Since we are not bound by automatic cut-off scores, the concern of the faculty for the validity of the GRE is not particularly high. They have in their own minds determined what it is worth, although their opinions vary.
>
> "Generally, the graduate dean believes verbal aptitude scores are very reflective of potential ability of master's students... Some areas (e.g., Psychology) agree, but many faculty do not and believe they ought to be eliminated."
>
> "Most everybody, if asked, will express reservations about the usefulness of GRE scores. However, the level of concern does not extend to the making of unsolicited proposals for [validity study]."

Reliance on subjective evaluations of predictive validity is implicit in such responses.

A number of respondents cited particular foci of concern, interest, or controversy in connection with the use of GRE scores in admissions including the following:

1) Are tests valid for predicting the performance of ethnic minorities or disadvantaged students? Older or part-time students? Foreign students?

2) Should uniform cutoff scores be used? Should weighted combinations of Verbal and Quantitative scores be employed? In general, what do GRE scores "mean" and what is the best way to consider scores in the admissions process?

Difficulties in conducting validity studies were mentioned by a number of respondents who cited such factors as criterion inadequacy (e.g., "no variation in graduate grades"), data collection problems (e.g., "no computerized student personal data summarization possible at present," "lack of clerical assistance"), and small samples.

Other respondents suggested that as standardized measures, GRE scores should serve as objective markers of student ability-levels (a) to help compensate for variations in the grading standards of undergraduate institutions, (b) to help monitor standards in the face of "grade inflation" at both the graduate and undergraduate levels, and/or (c) to help maintain and/or monitor "standards" among several disparate departments.

Complex problems are faced by graduate schools interested in developing systematic approaches to evaluation of the validity of admissions decisions generally, or in connection with the specification and maintenance of "standards." These problems are summarized rather succinctly in the comments of one dean:

> "Our concern is that we develop a balanced and far-ranging set of criteria for evaluating a widely disparate spread of applicants for widely disparate programs ranging from Anatomy and Anthropology to Theater Arts and Urban Planning.
>
> We undertook to require the GRE Aptitude as a uniform requirement for admission... (a) to give additional information on the increasing number of applicants from P/F, NR, Honors Exams schools and the like and, (b) to give us some counterweight to 'inflated grades'--or at least some additional standard of calibration.
>
> In Art (sculpture, ceramics, painting), the GRE Aptitude may have little application; in Art History, it may have high correlation. In Dance it may prove to have little use; in Economics it may have an important impact. And so I could go on throughout our 75 graduate degree-granting programs."

Not all graduate schools present such a wide-ranging set of programs, but the complexities reflected in the foregoing comment are to some degree characteristic of most graduate settings.

It would appear from many of the comments that in the graduate school community opinions and beliefs about GRE "validity" or "lack of validity" are strongly held despite the fact that studies designed to assess predictive validity in representative "use contexts" have not been made. Predictive validity frequently appears to have been perceived, erroneously, as an absolute test-quality rather than an expression of degree of relationship between two or more fallible measures (a predictor such as the GRE and a criterion such as the Graduate Grade Point Average), in particular samples. Generally speaking, both the comments and the findings regarding GRE validity study activity suggest that questions about GRE predictive validity are not perceived as recurring questions to which current answers frequently will be needed. In these circumstances, the goal of obtaining up-to-date empirical evidence regarding GRE-Validity is a challenging one.

Appendix B-3

Statement Regarding Confidentiality
and Study Procedures

COOPERATIVE GRE VALIDITY STUDIES
ONE-YEAR MODEL

Educational Testing Service
Princeton, NJ 08540

Begin B-3
Page 1 of 10 pages

To: PARTICIPANTS IN GRE VALIDITY STUDIES

Subject: Treatment of data on individuals for purposes of validation research

1. GRE validation research requires the linkage of information about the scores of individuals on GRE Aptitude and/or Advanced Tests and other predictive measures with information about their performance in graduate school.

   Provision of such information to organizations such as ETS is permissible under existing Federal legislation for purposes of "developing, validating or administering predictive tests [and for certain other designated purposes, if such studies are conducted in a manner as will not permit the personal identification of students and their parents by persons other than representatives of such organizations and such information will be destroyed when no longer needed for its original purposes]."

2. Participants in Cooperative Validity Studies are asked to submit information about the scores and performance records of students on validity study rosters. It should be noted that the names of students are not required to carry out the validity study analyses, and institutions may elect to eliminate names of students from the copies of rosters submitted to ETS. Some type of identification that will permit resolution of possible questions regarding missing, out-of-range, or improperly coded data should be substituted in such cases.

3. ETS procedures will be designed to protect the confidentiality of individual data in all cases. For institutions that elect, for any reason, to submit rosters containing names of students, the following procedures will be followed:

   a) After initial screening by project staff, for monitoring and editing purposes, data will be prepared for machine processing with numeric identification substituted for name identification.

   b) Original data rosters will be retained in a secure place for reference as required to resolve data-related questions that may arise during the course of the validity study process.

   c) Original data rosters will be retained under secure conditions no longer than is required to complete the sequence of activities involved in the validation research project and following completion of such activities the original rosters will be destroyed.

4. Names of individual students will in no way be involved in reports of validity study findings.

5. Names of institutions will not be identified with specific validity study findings in summary reports prepared for general distribution.

November, 1976

Appendix B-3

Statement Regarding Confidentiality and Study Procedures

COOPERATIVE GRE VALIDITY STUDIES
ONE-YEAR MODEL
Educational Testing Service
Princeton, NJ 08540

Begin B-3
Page 1 of 10 pages

To: PARTICIPANTS IN GRE VALIDITY STUDIES

Subject: Treatment of data on individuals for purposes of validation research

1. GRE validation research requires the linkage of information about the scores of individuals on GRE Aptitude and/or Advanced Tests and other predictive measures with information about their performance in graduate school.

   Provision of such information to organizations such as ETS is permissible under existing Federal legislation for purposes of "developing, validating or administering predictive tests [and for certain other designated purposes, if such studies are conducted in a manner as will not permit the personal identification of students and their parents by persons other than representatives of such organizations and such information will be destroyed when no longer needed for its original purposes]."

2. Participants in Cooperative Validity Studies are asked to submit information about the scores and performance records of students on validity study rosters. It should be noted that the names of students are not required to carry out the validity study analyses, and institutions may elect to eliminate names of students from the copies of rosters submitted to ETS. Some type of identification that will permit resolution of possible questions regarding missing, out-of-range, or improperly coded data should be substituted in such cases.

3. ETS procedures will be designed to protect the confidentiality of individual data in all cases. For institutions that elect, for any reason, to submit rosters containing names of students, the following procedures will be followed:

   a) After initial screening by project staff, for monitoring and editing purposes, data will be prepared for machine processing with numeric identification substituted for name identification.

   b) Original data rosters will be retained in a secure place for reference as required to resolve data-related questions that may arise during the course of the validity study process.

   c) Original data rosters will be retained under secure conditions no longer than is required to complete the sequence of activities involved in the validation research project and following completion of such activities the original rosters will be destroyed.

4. Names of individual students will in no way be involved in reports of validity study findings.

5. Names of institutions will not be identified with specific validity study findings in summary reports prepared for general distribution.

November, 1976

BASIC COOPERATIVE GRE VALIDITY STUDIES

Overview of Validity Study Data
Requirements and Procedures

A detailed set of instructions for participants in the basic GRE Validity Studies Project is attached. The purpose of this overview is to provide a brief description of requirements and procedures in order to permit an assessment of the types of data requirements and options involved for participants in studies. Institutions/departments are expected to provide data according to procedures outlined. ETS will analyze data and prepare a report without cost to participants.

The following applies to each participating department:

I. Focus of the study is on first-time graduate students enrolled in a degree program, and classifiable as full-time according to institutional/departmental definitions at time of entry into the department.

II. The sample to be studied consists of all such students who entered in Fall 1974 and 1975. At least 25 of these students should have GRE Aptitude and/or Advanced Test scores and at least one measure of performance in the department.

III. The validation period is to be the first year of study. For each student entering in Fall 1974, information regarding progress in the department as of Fall 1975 is to be provided; for those entering in Fall 1975, progress is to be encoded as of Fall 1976-77.

IV. A progress code is to be recorded for each student in the sample and at least one measure of performance should be recorded; several options are provided.

Thus, minimum requirements for participation in the basic validity studies are as follows:

A. List all first-time enrolled, full-time, degree-seeking students entering in 1974 and 1975.

B. Encode progress as of the beginning of the second year following admission for each student listed.

C. Record GRE Aptitude and/or Advanced Test scores for each student as available (at least 25 recommended).

D. Record at least one measure of performance for each student: e.g.,

1. First-year graduate grade point average or some function thereof
2. Performance in some critical course, course sequence, seminar, or common first-year project
3. Performance according to regular end-of-year departmental rating or examination procedures
4. Ad hoc ratings by faculty members according to one of two standard schedules or to some other schedule devised by a department.

THE FOREGOING REPRESENT CORE REQUIREMENTS. PLEASE SEE THE DETAILED OUTLINE FOR SUGGESTED OR RECOMMENDED CODES, RATING PROCEDURES, ETC.

Revised 10/76

BASIC COOPERATIVE GRE VALIDITY STUDIES

Instructions for Completing Validity Study Rosters: Definitions and Procedures

I. Definition of terms

A. First-time enrolled graduate student: No graduate study prior to enrollment in department/program. Some latitude discretionary with department in including individuals with limited previous graduate work if such individuals pursue first-year tasks similar to those of first-time enrolled students.

B. Full-time: Students classifiable as "full-time" graduate students according to institutional/departmental criteria.

C. Degree-seeking: Taking work creditable toward a graduate degree and considered by the institution/department to be prospective degree candidates. If a departmental sample includes both prospective master's- and doctoral-degree candidates, and if first-year tasks and/or evaluation procedures are not comparable for these two groups, the degree objective of a student should be coded as an optional data element (see instructions relating to Roster Columns 16 and 17, below).

II. Procedures for completing validity study rosters: one for each participating department. Instructions for each Roster Column are as follows:

ROSTER COLUMN 1. Identification: List all Fall 1974 and Fall 1975 entrants, respectively, classifiable as first-time enrolled, full-time, degree-seeking students at time of entry. [Name identification not required for validity study purposes, per se. See statement re treatment of data on individuals.] If students for whom English is not the native language are included they should be identified by coding as an optional data element (see instructions for Roster Columns 16 and 17, below).

ROSTER COLUMN 2. (Optional) Code for sex

Female = 1
Male = 2

ROSTER COLUMN 3. (Optional) Ethnic group code

1. American Indian, Eskimo, or Aleut
2. Black or Afro-American or Negro
3. Mexican-American or Chicano
4. Oriental or Asian-American
5. Puerto Rican
6. Other Hispanic or Latin American
7. White or Caucasian
8. Other

Revised 10/76

ROSTER COLUMN 4. Status code. For each student listed encode information regarding status as of the beginning of the second year following admission--i.e., for 1974 entrants, encode status as of Fall 1975; for 1975 entrants, encode status as of Fall 1976, as follows:

4 = Continuing in progress toward a degree in the department
3 = Not continuing in the department; completed first year in good standing
2 = Not continuing in the department; completed first year with one or more indications of marginal or substandard performance
1 = Not continuing in the department; did not complete first year

ROSTER COLUMNS (5), (6), and (7) are for recording GRE Aptitude and/or Advanced Test scores, as available, for each student listed. RECORD SCALED SCORES ONLY (e.g., 520, 780, etc.). Indicate name of Advanced Test field on the cover sheet (see instructions on sheet).

ROSTER COLUMN 5. Enter GRE Verbal scaled score in Column 5.

ROSTER COLUMN 6. Enter GRE Quantitative scaled score in Column 6.

ROSTER COLUMN 7. Enter GRE Advanced Test scaled score in Column 7. Identify Advanced Test field on cover sheet.

RECORD ONE OR MORE CRITERION SCORES FOR EACH STUDENT LISTED. Roster columns (8) through (14) are provided for recording one or more criterion scores, as available, for each student. Scores should reflect the assignment of a student to one of two or more ordered groups or categories in terms of level of performance (success, attainment, achievement) during the first year of study. AT LEAST ONE CRITERION MEASURE IS NEEDED TO CONDUCT A STUDY.

Several criterion measures are suggested, as follows:

ROSTER COLUMN 8. Overall Graduate Grades. (GPA, general) Performance as reflected in graduate grades, based on work completed during the first year (a grade point average or some function of grades earned such as, for example, "percent of grades that were A+ or A"; "all grades satisfactory = 1 versus one or more grades unsatisfactory or marginal = 0," etc.). DESCRIBE SCALES AND CODING PROCEDURES ON THE COVER SHEET.

ROSTER COLUMN 9. Grades in critical area. (Critical GPA) Performance in a critical course, course sequence, seminar, or project required of all or most first-year students, or normally completed by such students. Grade received in such a critical area, Pass = 1 versus Fail = 0, or other indication of standing should be reported in Column 9. DESCRIBE THE CRITICAL AREA, CODING, AND RELATED PROCEDURES ON THE COVER SHEET.

ROSTER COLUMN 10. Regular faculty ratings (Regular departmental evaluation) If regular faculty ratings of students constitute a part of the first year pattern, record rating in Column 10. DESCRIBE RATING SCALE AND PROCEDURES ON THE COVER SHEET.

ROSTER COLUMN 11. Regular departmental examinations. (End of year examination) Record score on end-of-year departmental examination in Column 11 (e.g., Pass = 1, fail = 0, or more refined score). DESCRIBE EXAMINATION, SCORING SYSTEM, ETC. ON COVER SHEET.

ROSTER COLUMNS (12), (13), and (14) ARE PROVIDED FOR AD HOC RATINGS FOR PURPOSES OF VALIDATION RESEARCH. Two standard rating schedules for faculty ratings of students are suggested for departments that do not employ regular rating procedures at the end of the first year. If neither of the suggested schedules is deemed to be appropriate, a department is encouraged to devise and apply a rating procedure that it considers to be appropriate. THE SUGGESTED SCHEDULES FOR FACULTY RATINGS ARE DESCRIBED ON A SEPARATE SHEET WHICH ALSO INCLUDES SUGGESTED PROCEDURES FOR OBTAINING RATINGS.

ROSTER COLUMN 12. Standard Rating, Schedule 1 (Rating relative to departmental standards and expectations) DESCRIBE PROCEDURES USED ON THE COVER SHEET. Only one rating to be reported for each student.

ROSTER COLUMN 13. Standard Rating, Schedule 2 (Rating in terms of potential for advanced study in a field based on rater-perception of general field demands or requirements) DESCRIBE PROCEDURES USED ON THE COVER SHEET. Only one rating to be reported for each student.

ROSTER COLUMN 14. Optional ad hoc faculty rating. DESCRIBE RATING SCHEDULE AND PROCEDURES ON THE COVER SHEET. Only one rating to be reported for each student.

NOTE REGARDING AD HOC RATINGS: These retrospective ratings should be based on observation of student performance in the department from time of entry through time of the rating (or time of student withdrawal from the department). Thus ratings for 1974 entrants typically will reflect observation over 2+ years of study while ratings for 1975 entrants will be based on observation over 1+ years of graduate study.

ROSTER COLUMNS (15), (16), and (17) ARE AVAILABLE FOR RECORDING ADDITIONAL DATA ELEMENTS.

ROSTER COLUMN 15. Undergraduate Grade Point Average (DESCRIBE AND IDENTIFY SCALE)

ROSTER COLUMN 16. OPTIONAL DATA ELEMENT (DESCRIBE AND IDENTIFY CODING OR SCALE)

ROSTER COLUMN 17. OPTIONAL DATA ELEMENT (DESCRIBE AND IDENTIFY CODING OR SCALE)

Examples of additional data elements and suggested coding are as follows:

a) Degree objective (Master's = 1, Doctorate = 2)

Use this code if department includes both master's and doctorate-seeking students and if first-year tasks and/or evaluation procedures are not comparable for the two groups.

b) Foreign student

1 = Foreign student (English not native language).

2 = Students for whom English is native language.

Use if foreign students are included on roster.

c) Year of bachelor's degree (record last two digits)

d) Year of birth (last two digits of birth year)

e) Undergraduate major same as graduate field

f) Received undergraduate degree from this institution

This institution = 1
Other institution = 0

g) Quality of undergraduate institution as judged by department (define procedures for establishing "quality")

High = 3

Medium = 2

Low or unknown = 1

h) Award status

Major fellowship or assistantship = 3
Other award or type of aid = 2
No award or financial aid = 1

i) IF ESTIMATES OF CANDIDATES' POTENTIAL ARE MADE ROUTINELY AS PART OF THE ADMISSIONS PROCESS, THE RATINGS, SCORES, OR CLASSIFICATIONS REFLECTING THOSE ESTIMATES COULD BE PROVIDED. THE TYPES OF INFORMATION USED IN ARRIVING AT THE ESTIMATES SHOULD BE DESCRIBED AS WELL AS THE PROCEDURES EMPLOYED.

For each departmental sample involved, a cover sheet should be completed. Space is provided for describing codes and identifying the data supplied. It is particularly important that each data element provided be described (e.g., nature of grading scale, rating procedures used, etc.).

If the supply of cover sheets and validity study roster forms is not sufficient please reproduce additional copies of the form. Additional copies will be forwarded upon request, however, if desired.

All materials when completed should be mailed as follows:

Cooperative Validity Studies Project
c/o Kenneth M. Wilson, R 208
Educational Testing Service
Princeton, NJ 08540

Call: 609-921-9000 for further information or clarification of procedures
Extension 2391

SUGGESTED RATING SCHEDULES AND PROCEDURES

A departmental faculty should select the rating schedule that it deems to be most consistent with its orientation to the assessment of student progress in the department. Only one rating should be reported for each student. Ratings should be based on observation of performance from entry to time of rating (or time of last official enrollment, if earlier).

Various procedures for obtaining ratings may be considered. For example:

(a) Ad hoc departmental committee to arrive at a "consensus" rating for each student listed; consultation with colleagues re cases not known to committee members or "difficult to assess" cases.

(b) Solicit ratings of listed individuals from departmental faculty members. Each faculty member to rate each student whose record is known. Ratings collected and collated for averaging. A minimum of two ratings required.

REGARDLESS OF THE PROCEDURES FOLLOWED, INDEX FINALLY DEVELOPED FOR EACH STUDENT SHOULD BE ENTERED IN COLUMN 12, 13, or 14 of the Validity Study Roster, depending upon use of Schedule 1, Schedule 2, or a schedule devised by the participating department.

PROCEDURES EMPLOYED IN DEVELOPING THE RATINGS SHOULD BE DESCRIBED.

Standard Rating Schedules

Schedule 1. Rating relative to departmental expectations and standards

Taking into account departmental expectations and standards, how would you characterize this student's record in the department?

4 = Distinguished

3 = Good to Strong

2 = Adequate to Adequate plus

1 = Unsatisfactory to Marginally Adequate

Schedule 2. Rating in terms of potential for advanced study in a field

Based on your observation of this student's performance how would you characterize his or her potential for advanced study in this field, given your perception of general field demands and requirements?

4 = Outstanding performer; definitely qualified for doctoral study

3 = Definitely master's caliber; probably capable of acceptable doctoral study

2 = Adequate to adequate plus at the master's level; would not encourage doctoral study

1 = Unacceptable or only marginally acceptable for graduate study at the master's level

Revised 10/76

ETS USE

GRE number______ Departmental code______

Page 1 of 2 pages

VALIDITY STUDY COVER SHEET

Institution______________________________________ Department____________________
(Name) (City/State)

Highest degree offered within department: (1)_______Master's (2)_______Doctorate

This sample of first-time enrolled, full-time students includes (check applicable statements):

______(1) Only prospective candidates for a master's degree

______(2) Only prospective candidates for a doctoral degree

Both prospective master's and prospective doctoral degree candidates

______(3) for whom first-year programs and evaluation procedures are comparable.

______(4) for whom first-year programs and/or evaluation procedures are not comparable. [If (4), code degree objective of each student in Optional Data column, as indicated in instructions. Elaborate on reverse side of this form.]

PLEASE CHECK (✓) IN A BOX, BELOW, TO INDICATE THAT THE DATA ELEMENT DESIGNATED IS PROVIDED FOR THIS SAMPLE.

	Optional coding		Status code	GRE scores (scaled)			First year grades		Regular departmental eval.		Ad hoc departmental rating			Optional additional data		
	Sex **	Ethnic group **	Status	V **	Q **	Adv.	GPA gen.	Critical GPA	Rating	Exam score	Sch 1	Sch 2	Sch 3	UGPA	Other	Other
ROSTER COLUMN	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)

** No explanation or further elaboration required if suggested validity study procedures have been followed. For noting exceptions to study procedures for these elements and for describing other data provided, use the spaces below. Please identify data elements described by column number.

ROSTER COLUMN	Description of data provided, including description of procedures (e.g., scales of grade point averages, nature of regular rating procedures or departmental examinations, methods of obtaining ad hoc ratings, etc.)
	OVER

GRE # ________ (ETS Use)
Institution/Department ______________________________ Page________ of ________pages

GRE VALIDITY
STUDY ROSTER.......

		Optional Coding		Status Code	GRE scaled scores			First-year grades		Regular departmental eval		Ad hoc departmental rating			Optional additional data		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	Identification	Sex	Ethnic group	Status	V	Q	Adv	GPA gen	GPA Crit	Rating	Exam score	Sched 1	Sched 2	Sched 3	UGPA	Option	Option
1.																	
2.																	
3.																	
4.																	
5.																	
6.																	
7.																	
8.																	
9.																	
10.																	
11.																	
12.																	
13.																	
14.																	
15.																	
16.																	
17.																	
18.																	
19.																	
20.																	

Appendix B-4

BRIEF DESCRIPTION OF SELECTED STUDIES INVOLVING A TWO-YEAR, SINGLE-COHORT STUDY MODEL

An effort was made during the course of the project to develop a multi-institutional, multi-departmental study calling for the concurrent participation of the same set of departments from several similar institutions. Several graduate schools with established arrangements for sharing data on admissions-related questions expressed an interest in participating in a study, based on a two-year study model, involving collection of data on first-time graduate students who entered Ph.D. oriented programs in the Departments of English, Romance Languages, Philosophy, Economics, Geology and Geophysical Sciences, and Physics, respectively, in Fall 1974

It was thought that participation of a common set of departments from several similar institutions would permit (a) the collection of comparable criterion data, beyond grade averages, such as scores on comprehensive examinations and/or standard faculty ratings and (b) analyses based on pooled data for the respective departmental samples. Information was provided by several of the departments regarding their examination practices.

There was substantial variation in the timing, scope, and coverage of the examinations, a set of factors that miliated against use of departmental examinations as a common criterion. Also, the goal of obtaining systematic faculty ratings, according to a standard schedule for students in the respective departments at each interested school, was not realized.

These considerations, and the inability of several of the originally interested schools to provide data, effectively precluded development of the study along the multi-institutional, multi-departmental lines originally proposed. However, four institutions provided data for samples from five to eight departments. The study called for data for only one entering cohort. In consequence, sample size was unusually small, as indicated below:

Institution A.	N's ranged from six to 24 per department over six departments.
Institution B.	N's ranged from five to 19 in five departments.
Institution C.	N's ranged from four to 35 in eight departments, with median N = 14.
Institution D.	N's ranged from six to 66 over six departments, with median N = 19.

No single criterion was common to all schools and departments;

and one or more predictor observations were missing for one or more students in most of the samples. Individualized institutional reports were prepared for each of the four schools. However, lack of uniformity in data militated against the summarization of data across schools. Results within the respective schools were consistent with the general proposition that GRE scores should tend to be positively related to performance in graduate study.

Planning and implementing a study calling for the concurrent participation of a designated set of departments from each of several institutions clearly posed considerably more complex problems than those involved in planning and implementing the basic studies that called for institutions to submit data for one or more departments selected on the basis of local interests and priorities.

Data collection materials: Two-year model

Page 1 of 10 pages

GRE COOPERATIVE VALIDITY STUDY*

Instructions for Completing Validity Study Rosters: Definitions and Procedures

I. Definitions

A. First-time Enrolled Graduate Student: No graduate study prior to enrollment in the department. Some latitude permissible for the inclusion of individuals with limited previous graduate study.

B. Full-time: Students classifiable as "full-time" students according to institutional-departmental criteria at time of entry.

C. Degree-seeking: At time of admission, was considered by the department to be a prospective doctoral-degree candidate.

II. Procedures for completing validity study rosters - A study roster should be completed for each department. Instructions for completing the study roster are as follows:

ROSTER COLUMN (1) Identification - List all fall 1974 entrants meeting the definitions outlined above. Note that name identification is not required for purposes of the validity study and names of students may be deleted from any rosters prior to their transmittal to ETS.

If noncitizens of the U.S. whose lack of fluency in English may have constituted a handicap in completing GRE requirements are included they should be identified by special coding in Column 3 (see instruction for that column, below).

ROSTER COLUMN (2) Sex (optional coding)

1 = Female
2 = Male

ROSTER COLUMN (3) Ethnic Group Code (optional, but desirable for validation research)

U.S. Citizens

1 = American Indian, Eskimo, or Aleut
2 = Afro-American, Negro, or Black
3 = Mexican-American or Chicano
4 = Oriental or Asian-American
5 = Puerto Rican
6 = Other Hispanic or Latin-American
7 = Caucasian or White
8 = Other (disadvantaged) minority not classifiable above

Non-U.S. Citizens

9 = "Foreign Student" (circle the code if English is native language)

---

*Two-Year Model (Ph.D.-oriented programs)
May 1977

ROSTER COLUMN (4) For each student listed, encode information regarding progress in the department as of spring 1977, according to the following classification:

Code

6 = Continuing the department; on or ahead of schedule in meeting applicable degree requirements

5 = Continuing in the department; somewhat behind schedule in meeting applicable degree requirements (no discontinuities in attendance—not counting summer sessions)

4 = Continuing in the department; some discontinuities in attendance and delays in meeting applicable requirements

3 = Not continuing in the department; cumulative record at time of withdrawal was satisfactory

2 = Not continuing in the department; cumulative record at time of withdrawal included some indications of sub-standard or marginal performance

1 = Not classifiable above (Describe on the cover sheet the patterns included in this category.)

ROSTER COLUMNS (5), (6), and (7) ARE FOR RECORDING GRE APTITUDE AND ADVANCED TEST SCALED SCORES, AS AVAILABLE, FOR EACH STUDENT. RECORD THE SCALED SCORES ONLY (E.G., 520, 780, ETC.). INDICATE ON THE COVER SHEET THE ADVANCED TEST FIELD(S) REPRESENTED. IF MORE THAN ONE FIELD, NOTE ALL FIELDS ON THE COVER SHEET. WHEN RECORDING ADVANCED TEST SCALED SCORES INDICATE EXCEPTIONS TO THE MAJORITY PATTERN BY WRITING IN FIELD ABBREVIATION (E.G., FIRST TWO OR THREE LETTERS, OR MORE AS REQUIRED TO IDENTIFY FIELD) OVER THE SCALED SCORE IN COLUMN (7).

ROSTER COLUMN (5) GRE Verbal Scaled Score

ROSTER COLUMN (6) GRE Quantitative Scaled Score

ROSTER COLUMN (7) GRE Advanced Test - Remember to write in field name above scaled score entry for all exceptions to the majority Advanced Test field.

ROSTER COLUMN (8) Undergraduate Grade Point Average (UGPA) (optional, but desirable if available) - Enter the UGPA in Column (8) as normally computed and used in the admissions process. The scale employed should be described on the cover sheet.

ROSTER COLUMN (9) Admissions Rating or Ranking (optional, but desirable if available for all or most students) - Enter any systematic ranking or rating reflecting an admissions-related assessment of relative potential or promise. If for example admitted applicants were classified on the basis of their admissions credentials, the "ranking" involved should be entered for analysis in relation to the criterion variables reported. PROCEDURES USED IN ARRIVING AT THE ADMISSIONS RATING SHOULD BE DESCRIBED ON THE COVER SHEET.

ROSTER COLUMNS (10) THROUGH (13) ARE TO BE USED FOR RECORDING UP TO FOUR CRITERION "SCORES" FOR EACH STUDENT. EACH PARTICIPATING DEPARTMENT IS ENCOURAGED TO PROVIDE A CUMULATIVE GRADUATE GRADE POINT AVERAGE AND AN AVERAGE OR "CONSENSUS" RATING OF STUDENT PERFORMANCE ACCORDING TO A SCHEDULE OUTLINED IN AN ATTACHMENT. AT LEAST ONE CRITERION MEASURE IS NEEDED TO CONDUCT A STUDY. EACH DEPARTMENT INTERESTED IN DOING SO MAY REPORT "SCORES" ON QUALIFYING, COMPREHENSIVE, GENERAL, OR PROFICIENCY EXAMINATIONS NORMALLY SCHEDULED FOR THE FIRST AND/OR SECOND YEAR OF GRADUATE STUDY. IF FACULTY RATINGS OF STUDENTS ARE REGULARLY MADE AFTER ONE OR TWO YEARS OF STUDY, THOSE RATINGS MAY BE REPORTED.

ROSTER COLUMN (10)	Cumulative Graduate Grade Point Average (GPACUM) - Record the grade point average based on course work completed during the first two years of graduate study, or all work completed prior to a student's withdrawal from the department, if applicable. DESCRIBE GRADE SCALE AND AVERAGING RULES ON THE COVER SHEET.
ROSTER COLUMN (11)	Ad Hoc Rating of Student Performance Relative to Departmental Expectations and Standards - If a department elects to develop *ad hoc* ratings for purposes of validity study, the ratings should be based on observations of performance in the department from time of entry to time of rating for currently enrolled students, and from time of entry to time of withdrawal for others. A standard rating schedule is attached. Use of the schedule outlined in the attachment is encouraged. However, if some other *ad hoc* procedure is deemed more appropriate, a department should feel free to use that procedure.
ROSTER COLUMN (12)	"Score" on Critical Examination (qualifying, general, proficiency) Enter here a summary score reflecting a student's performance on the first critical examination that members of an entering cohort may be expected to have attempted during the first and/or second year of graduate study. To be considered "critical" the examination(s) involved must be met by all aspirants to a doctoral degree. At least Pass/Fail and preferably a more refined gradation of performance should be reported. The nature of the examination(s) involved should be described on the cover sheet along with a description of the scoring and the scores reported.
ROSTER COLUMN (13)	Regular Faculty Ratings of Student Performance - Record in this column ratings of student performance that may have been made at the end of the first or second year of study, as part of the normal or regular pattern of departmental procedures. The timing of the ratings (e.g., end of first year) and the procedures employed as well as the place of the ratings in the total pattern of departmental requirements should be described.

ROSTER COLUMNS (14) THROUGH (17) MAY BE USED TO RECORD ANY OTHER PREDICTOR OR CRITERION DATA OF INTEREST TO A DEPARTMENT. EXAMPLES OF SUCH DATA, AND SUGGESTED CODING PATTERNS ARE LISTED BELOW:

- a) Year of bachelor's degree (last two digits)
- b) Year of birth (last two digits)
- c) Undergraduate major same as graduate field = 1, other = 0
- d) Received undergraduate degree from this institution = 1, other = 0
- e) Received undergraduate degree from highly selective institutions (e.g., DWARFS = 1, other = 0)
- f) Award status
  - 4 = Holds or has held major research fellowship or assistantship
  - 3 = Holds or has held major nonresearch fellowship or assistantship
  - 2 = Holds or has held meaningful but not major fellowship or assistantship
  - 1 = Holds or has held no type of fellowship or assistantship

IT IS IMPORTANT TO DESCRIBE EACH OPTIONAL DATA ELEMENT FULLY.

For each departmental sample, after the validity study roster is completed a "cover sheet" (special form) should be prepared. Space is provided on the cover sheet for describing all "nonstandard" codes and identifying the data supplied. It is particularly important that each data element provided be described (e.g., scales for GPA variables, procedures used in ratings, codes for each categorical variable reported, etc.).

If the supply of cover sheets and validity study rosters is not sufficient, additional copies may be reproduced locally.

All materials when completed should be mailed as follows:

Cooperative Validity Studies Project
c/o Kenneth M. Wilson, R-208
Educational Testing Service
Princeton, New Jersey 08540

Call: 609-921-9000 for further information
or clarification of procedures

Attachment

GRE COOPERATIVE VALIDITY STUDY

Instructions for Completing Ratings: Rating Procedures

Recommended Standard Rating Schedule

If a department elects to develop *ad hoc* ratings for purposes of GRE validation research, it is recommended that the ratings be developed according to the following schedule:

> Taking into account departmental expectations and standards, how would you characterize the student's record in the department?
>
> 4 = Distinguished
> 3 = Good to Strong
> 2 = Adequate to Adequate Plus
> 1 = Unsatisfactory to Marginally Adequate

Ratings may be obtained by various means. For example:

a) By achieving at least two independent ratings for each student which can then be averaged—e.g., have all faculty members rate each student known to them; obtain rating from a student's advisor plus one additional faculty member, etc.
b) By having an *ad hoc* departmental committee develop a "consensus" rating for each student.

Regardless of the procedures followed, the average rating or the consensus rating developed for each student should be entered in Column 11 of the validity study roster.

Procedures employed in developing the ratings reported should be described on the cover sheet.

NOTE: Members of the entering cohort who are no longer enrolled in the department should be rated on the basis of performance during their period of enrollment if, in the judgment of faculty members involved, a ratable pattern of attainment was established during that period of enrollment.

For ETS Use Only
GRE Number ______________
Departmental Code ________

Page 1 of 4

VALIDITY STUDY COVER SHEET

Institution ____________________________________ Department ______________
(Name) (City/State)

Highest degree offered within department: (1)______Master's (2)______Doctorate

This sample of first-time enrolled, full-time students includes (check applicable statements):

______(1) Only prospective candidates for a master's degree

______(2) Only prospective candidates for a doctoral degree

Both prospective master's and prospective doctoral degree candidates:

______(3) for whom first-year programs and evaluation procedures are comparable, or

______(4) for whom first-year programs and/or evaluation procedures are not comparable. (If (4), code degree objective of each student in Optional Data column, 1 = Master's, 2 = Doctorate.)

PLEASE CHECK (✓) IN A BOX BELOW TO INDICATE THAT THE DATA ELEMENT DESIGNATED IS PROVIDED FOR THIS SAMPLE.

(2) Sex	(3) Ethnic Code	(4) Progress	(5) GRE-VERB	(6) GRE-QUANT	(7) GRE-ADV	(8) UGPA	(9) Admis. Rate	(10) GPA CUM	(11) Ad hoc Rate	(12) Crit Exam	(13) Reg Rate	(14) Option	(15) Option	(16) Option	(17) Option

NOTE: In the spaces below, provide information required to interpret each of the data elements checked above. Where standard coding is provided, only exceptions to that coding need be described.

ROSTER
COLUMN

(2) Sex (Female = 1, Male = 2) ______________________________

(3) Ethnic Group Code (per instructions, standard) ______________________

(4) Progress Code (per instructions, standard) ________________________

(5) GRE Verbal Scaled Score (standard) ____________________________

(6) GRE Quantitative Scaled Score (standard) _________________________

ROSTER
COLUMN

(7) GRE Advanced Test Score - Please list the test titles (fields) for which test scores are available. Also show the code or abbreviation with which the test is identified on the validity study roster, per se. See instructions for Column (7).

Code

Majority (or only) field ______________________ (a)______

Second field ______________________ (b)______

Third field ______________________ (c)______

Other ______________________ (d)______

(8) Undergraduate Grade Point Average - Describe (e.g., overall cumulative, upper division cumulative, major field only, scale employed, etc.).

____

____

____

(9) Admissions Rating or Ranking - Describe here the rank, rating, or composite score (e.g., 2 V+Q + 25 UGPA) reported.

____

____

____

(10) Cumulative Graduate Grade Point Average - Describe scale, method of computation (e.g., how hours in denominator are accumulated).

____

____

____

(11) Ad Hoc Departmental Rating (standard) - Describe how rating was determined (e.g., by averaging two or more independent ratings, by consensus procedures, typical number of ratings involved in average, etc.) USE SEPARATE SHEET IF SPACE HERE IS NOT SUFFICIENT.

____

____

____

____

____

ROSTER
COLUMN

(12) Critical Examination "Score" - In these spaces provide a brief description of the examination(s) involved, and a translation of scores reported, if numeric (e.g., High Pass = 3, Pass = 2, Qualified Pass = 1, Fail = 0). Use a separate sheet to describe the examination process more fully. NOTE: IF A DEPARTMENT HAS PREVIOUSLY PROVIDED A DETAILED DESCRIPTION OF THE EXAMINATION PROCESS, THIS ITEM NEED NOT BE COMPLETED.

(13) Regular Faculty Rating of Students - Describe here timing and purpose of ratings, procedures used, etc. Provide information required to translate numeric coding.

(14) Optional Data Element (describe fully)

(15) Optional Data Element (describe fully)

(16) Optional Data Element (describe fully)

(17) Optional Data Element (describe fully)

Information About the Admissions Process and Student Progress

(optional, but highly desirable if available)

Please answer the following questions based on studies that may have been done in the department or "best estimates" if no studies have been done or if statistics related to the question have not been regularly maintained. The questions pertain to individuals like those included in the present sample (e.g., not "transfer" graduate students).

1. What percentage of applicants typically is admitted? ____________

2. What percentage of admissions offers typically is accepted? ________________

SHOULD TOTAL 100%

3a. Give your best estimate of the percentage of first-time graduate students in a given cohort likely to complete Ph.D. requirements in the department. ________

3b. What percentage of an entering cohort is likely to fail to qualify for Ph.D. candidacy (e.g., withdraw with unsatisfactory record in course work, failure on qualifying examinations, failure to attempt specific requirements on schedule)? ________________________________

3b. What percentage of an entering cohort is likely to leave the department without a degree, but with a basically satisfactory record of performance (e.g., no "failure" to meet explicit requirement)? ________________________

4. What is the best estimate of median years to the Ph.D., matriculation to degree conferral, for first-time entrants like those included in the sample, who complete degree requirements in the department? ________________________

__________ CARD NUMBER
__________ COLLEGE NUMBER
__________ CLASS
__________ OF __________ PAGES

VALIDITY STUDY ROSTER

Institution/Department ______________________________

Optional Data

	(1) IDENTIFICATION	(2) Sex	(3) Ethnic Code	(4) Progress	(5) GRE-VERB	(6) GRE-QUANT	(7) GRE-ADV	(8) UGPA	(9) Admis. Rate	(10) GPA CUM	(11) Ad hoc Rate	(12) Crit Exam	(13) Reg Rate	(14) Option	(15) Option	(16) Option	(17) Option
1.																	
2.																	
3.																	
4.																	
5.																	
6.																	
7.																	
8.																	
9.																	
10.																	
11.																	
12.																	
13.																	
14.																	
15.																	
16.																	
17.																	
18.																	
19.																	
20.																	

Appendix C

C-1. Illustrative study report

C-2. Tabular summary of selected study findings

Appendix C-1

Illustrative Study Report

GRE COOPERATIVE VALIDITY STUDIES REPORT

Educational Testing Service
Princeton, NJ 08541

Validity Study Report

for the Departments of

BOTANY/MICROBIOLOGY
GEOLOGY/GEOPHYSICS

in

Graduate School A

May 1979

Sponsored by the Graduate Record Examinations Board

Appendix C-1

INTRODUCTION

This report has been prepared as part of the Cooperative Validity Studies Project being conducted by Educational Testing Service for the Graduate Record Examinations Board (GRE). The project is designed to help graduate schools and departments generate up-to-date evidence regarding the levels and patterns of relationships between GRE scores (and other data used in admissions, such as the Undergraduate Grade Point Average or UGPA) and one or more limited but relevant measures of performance in graduate school during the first year of study (e.g., criterion measures such as a Graduate GPA, faculty ratings, or departmental examinations), in recently enrolled cohorts (1974 and 1975 entrants) of first-time enrolled, full-time, degree-seeking students.

GRE validity study findings for the departmental sample(s) for which your institution provided data are presented herein. Information is provided regarding the level, patterning, and distribution of scores for the sample(s) on each of the variables for which data were submitted. More specifically, measures of central tendency (the arithmetic average or mean) and variability of scores (the standard deviation) are reported.

Evidence regarding the relationship between each admissions variable, or predictor (such as the GRE-Verbal, GRE-Quantitative, or UGPA), and each performance or criterion variable (such as a Graduate GPA) is presented in terms of the coefficient of correlation, a generally familiar index of association or covariation between variables. The size of a coefficient indicates the degree or closeness of association between two variables on a scale ranging from .00 (indicating no relationship at all) through ± 1.00 (indicating either a perfect positive or a perfect negative relationship). A positive sign indicates that higher standing on a predictor tends to be associated with higher standing on a criterion variable, whereas a negative sign indicates that higher standing on a predictor tends to be associated with lower standing on a criterion. When used to express the relationship between predictor and criterion measures in admissions settings, observed correlation coefficients (known as validity coefficients) are almost always positive. In GRE-validity studies, for example, coefficients between .20 and .30 are typical for individual predictors (cf. Table 1, p. 5).

Some Limiting Considerations

As you examine this report, it is quite important to keep in mind the limited scope of the validity assessment that is being undertaken in these Cooperative Studies. They focus primarily on first-year performance criteria and they are based on quite small samples in most instances. Small-sample analysis in any context calls for care in the interpretation of results, of course. Added interpretational complications arise in small sample GRE validity studies (a) since the distributions of potential predictors (e.g., GRE Aptitude Test scores and UGPA) tend to be restricted in range because they were "used" in the recruitment/selection process and (b) since local "validity study norms," derivable only through periodic replication of studies, may be either lacking or inadequately developed in many departmental settings.

In evaluating validity study findings generally it should be recognized that "predictive validity" is always relative--validity is not an absolute quality of a GRE test or any other admissions variable. A validity coefficient is simply an index of the relationship between a predictor and a given criterion, both of which are less than completely reliable measures, in a given prediction context and sample. The validity coefficients reported herein, therefore, apply only to the specific criterion measure(s) employed and the sample(s) or group(s) in which they were determined. Results should not be generalized to other criteria or groups.

Questions regarding the predictive validity of GRE scores for particular subgroups must be addressed in appropriate samples of the various subgroups of interest--e.g., "older" students, "transfer" students, or "minority" students. In the present series of studies, the focus is on "first-time enrolled, full-time, degree-seeking students." Determination of validity coefficients in more refined subgroups is not feasible due to the small size of the sample(s).

It is important to recognize in passing that the first-year criteria under consideration in this study (such as Graduate GPA or end-of-year ratings) are themselves subject to empirical evaluation as potential predictors of longer-term criteria of "success" in graduate programs (e.g., what is the observed validity of first-year Graduate GPA for predicting graduation versus nongraduation, or performance on Ph.D. qualifying examinations?).

Normative Perspective and Interpretational Guidelines

Evidence from validity studies that have been conducted in a variety of educational settings, extensively in undergraduate and law schools and less extensively at the graduate level, as well as evidence of the positively interrelated organization of human abilities, suggests that validity coefficients for academic predictors such as the GRE and UGPA and academic criteria such as Graduate GPA or faculty ratings should tend to be positive. In essence, it is reasonable to assume that individuals with "better qualifications" (as reflected in their past academic record and their scores on verbal and quantitative ability measures, for example) should tend to be somewhat "better students" (as reflected in faculty assessments of their work in graduate courses, seminars, and the like, for example).

If negative coefficents are observed in validity studies, they usually are small and may be explained as falling within the normal range of expected variation due to sampling error around a "characteristic" or "population" value that is low but probably positive. Negative coefficients for academic predictors and academic criteria are, therefore, properly perceived as theoretically anomalous. When observed, they indicate the need for further exploration and analysis designed to illuminate the particular circumstances involved.

It has been established that the size of validity coefficients tends to vary inversely with the degree of restriction of range of talent in samples being studied. The interpretation of observed validity coefficients for GRE scores or UGPA is especially complicated by the fact that graduate students, generally, represent a highly select group with respect to academic ability and past performance. In departmental samples such as those involved in validity studies, further restriction of range on these variables is introduced either directly (when GRE and UGPA have been used in selection) or indirectly (when other related variables may have been used). Restriction in range on one or more of the predictor variables under consideration makes it difficult to obtain a clear assessment of the actual "value" of the predictors involved since observed validity coefficients tend to be lower than would be the case if a "full range of talent" (e.g., a group representative of all college seniors) could be included in departmental samples.*

Restriction in the range of criterion values also complicates the interpretational outlook. If criterion values, such as Graduate GPA, vary only over a very limited range (e.g., A or B) differences in student performance may not be measured reliably and this also tends to lead to underestimation of the overall utility of a predictor.

Some evidence regarding the characteristic levels of validity coefficients that have been obtained in representative studies involving GRE and UGPA and a Graduate GPA criterion is provided in Table 1 which shows median coefficients from a number of studies, as recently summarized by Willingham (1974).** Also shown are data from undergraduate validity studies involving a GRE-comparable measure, namely, the CEEB Scholastic Aptitude Test (Verbal or SAT-V and Mathematical or SAT-M) and college freshman-year GPA (Schrader, 1971). These latter data indicate how validity coefficients tend to be lower in samples that are highly selected on verbal ability than in samples that are more representative with respect to verbal ability.

In the graduate school studies, median validities for the GRE Aptitude Test components based on studies involving Graduate GPA criteria in samples from a variety of disciplines were slightly higher than .20, those for the GRE Advanced Test or the UGPA, alone, were about .30, while the best-weighted (i.e., multiple regression based) combinations of GRE Aptitude scores and UGPA yielded validity (multiple correlation) coefficients averaging around .45. Note that these coefficients are similar in pattern and level to those observed in undergraduate settings in which samples were highly restricted with respect to SAT Verbal scores (i.e., high mean and small standard deviation).

---

*In recent years, GRE-Verbal scores for candidates nationally have had standard deviations of approximately 125, and the standard deviations of GRE-Quantitative scores have been approximately 135. In departmental samples such as those involved in the present studies, standard deviations of 75 to 90 on one or both these variables are not uncommon, indicating that the range of ability available for study is considerably less than that in the total group of individuals taking the GRE Tests nationally.

**Warren W. Willingham, Predicting success in graduate education, Science, 183, pp. 273-278. This is a brief but comprehensive overview not only of representative GRE validity study findings during 1952-72 but also of basic validity-study concepts, problems, and issues.

Table 1

Median Validity Coefficients and Range of Coefficients in Studies of Comparable Predictors and Grade Average Criteria in Graduate and Undergraduate Settings

Predictors as appropriate to level of study*	GRE studies in graduate school settings	CEEB SAT-validity studies in undergraduate samples which were		
		High & homogeneous on SAT-V**	Homogenous on SAT-V***	Representative SAT-V scores*
	Mdn.	Mdn. (range)#	Mdn. (range)	Mdn. (range)
GRE-Verbal (Grad.)	.24 (46)##			
SAT-Verbal (U.G.)		.22 (.11 to .44)	.31 (.15 to .46)	.39 (.26 to .54)
GRE-Quant. (Grad.)	.23 (43)			
SAT-Math (U.G.)		.24 (-.01 to .46)	.27 (.11 to .40)	.33 (.20 to .48)
GRE Advanced (Grad.)	.30 (25)			
Undergrad. GPA	.31 (26)			
High School Record		.40 (.32 to .57)	.44 (.26 to .59)	.55 (.33 to .67)
GRE + UGPA (Grad.)	.45 (24)			
SAT + HSR (U.G.)		.46 (.35 to .61)	.52 (.34 to .66)	.62 (.46 to .73)

Note: Graduate school data are from Willingham, W.W., Predicting success in graduate education, Science, 183, 1974, 273-278, Table 1. Undergraduate validity data are from Schrader, W. B., The predictive validity of College Board admissions tests, in Angoff, W. H. (Ed.), The College Board admissions testing program. (Princeton, NJ: College Entrance Examination Board, 1971), pp. 117-146.

*The coefficients in column 1 of the table reflect validities for GRE-Verbal, GRE-Quantitative, GRE Advanced, Undergraduate GPA, and GRE-UGPA composites, respectively, in graduate school samples. All the remaining coefficients reflect validities for SAT-V, SAT-M, the high school record, and SAT-HSR composites, respectively, in samples of college freshmen.

**Studies in 18 samples of having an SAT-Verbal mean above 600 and standard deviation of 65 or less (undergraduate freshmen).

***Studies in 95 samples of undergraduate men and women freshmen having SAT-Verbal standard deviation of less than 75. Median values reported separately for men and women by Schrader have been averaged for presentation in this table.

#Approximately 80 percent of the obtained coefficients were within the range specified.

##Number of coefficients upon which each median is based. Studies are summarized without regard to the field of study involved.

Table 2

Variation in Validity Coefficients for Selected Predictors in Studies Replicated Across and Within Undergraduate Colleges

College/Class	Typical sample size	SAT-V	SAT-M	Rank	Achievement test aver.
College A, 1968	(352)	.16	.13	.27	.32
1969		.37	.23	.37	.37
1970		.31	.28	.32	.52
1972		.30	.20	.28	.35
College B, 1968	(266)	.20	.21	.43	.42
1969		.29	.11	.39	.37
1970		.27	.05	.31	.22
1972		.21	-.06	.20	.24
College F, 1968	(33)	.23	.13	.42	.31
1970		.25	-.05	.21	.43
1972		.28	.13	.39	.24
College J, 1968	(62)	.27	.49	.61	.43
1969		.20	.33	.47	.46
1970		.48	.54	.54	.56

Source of data: Kenneth M. Wilson, The contribution of measures of aptitude (SAT) and achievement (CEEB Achievement Average), respectively, in forecasting college grades in several liberal arts colleges, Research Bulletin 74-36, Educational Testing Service, 1974, Table 4.

These findings suggest that prediction of Graduate GPA from GRE scores or Undergraduate GPA might be accomplished at about the same level of "accuracy" as that involved in predicting college freshman-year GPA using comparable predictors in "high ability-low variablity" undergraduate settings but at a lower level of accuracy than that found in more representative undergraduate settings and samples.

Some indication of the range of coefficients obtained in replications of validity studies across colleges is provided in the "range" data shown in Table 1. Table 2 provides illustrative data showing how coefficients may vary under conditions of replication within a college as well as across colleges. The coefficients shown are for samples of undergraduate women in successive classes in several selective colleges (Wilson, 1974). Among other things, these data suggest (a) that "small sample" studies do not necessarily yield less stable or interpretable results even though the potential for marked variability due solely to sampling error is greater in such samples, (b) that regardless of sample size, inferences regarding "the validity" of a given predictor or the relative validity of several predictors within a given prediction context should not be drawn on the basis of one validity study, and (c) that the levels and patterns of validity coefficients that obtain for one educational setting do not necessarily hold for another.

Questions of "Weighting"

In validation research generally, it has been found that an appropriately weighted combination of two or more available predictors frequently yields higher validity coefficients than any of the predictors considered separately. This pattern is suggested, for example, by the GRE validity study findings summarized in Table 1. Combining GRE Aptitude scores and the UGPA tends to result in improved validity.

In large-sample validity studies, questions regarding the most effective weighting of two or more predictors with respect to a given criterion traditionally have been addressed by application of appropriate multivariate statistical methods, principally multiple correlation and regression analysis. One typical outcome of these studies has been the development of "equations" for obtaining a "predicted criterion score" for each individual. Weights indicating the relative contribution of two or more predictors in such equations are specified in the analysis. In small samples, however, multivariate procedures have more limited operational utility due to the fact that the results obtained (the weights derived) may reflect too closely the possibly idiosyncratic patterns in a small sample ("overfitting the data") and thus vary markedly in subsequent small samples.

Given the typically small size of the sample(s) under consideration in these GRE Cooperative Studies, and the additional limitations imposed by the fact that particular observations may be lacking in some instances (e.g., some individuals in the sample may not have GRE scores, or a criterion measure), any application of multivariate analysis that may be reported herein is intended only to facilitate limited consideration of certain of the principles, advantages, and persistent problems involved in developing reliable information about the relative contribution of predictors and in combining two or more predictors in admissions settings.

All Composites of GRE Scores Require Empirical Validation

It is particularly important to call attention to the potential hazards involved in using any "intuitively appealing" procedure for combining GRE scores that has not been carefully evaluated empirically. Use of a GRE-Total (a simple additive combination of the GRE-Verbal and GRE-Quantitative scores) constitutes a highly relevant "real life" example of a procedure that should be avoided in the absence of empirical evidence that GRE-Total is more valid than either of its components. In certain circumstances not infrequently encountered in practice GRE-Total actually is less valid than GRE-Verbal or GRE-Quantitative only, whichever has the higher validity.

To illustrate certain approaches to weighting, as well as some of the problems involved, consideration is given to the validity of various composites of available predictors. Again it is important to stress the fact that the purpose of doing so is primarily didactic.

A Brief Interpretational Rationale

Assessment of the "meaning" or "significance" of validity coefficients obtained in first-time validity studies in small departmental samples obviously should be undertaken cautiously.* Inferences regarding the relative importance or validity of GRE-V, GRE-Q, GRE-Advanced, UGPA and/or other predictors, should be drawn only tentatively and entertained as hypotheses calling for further verification. Attention should be focussed on discernible trends, consistencies, and inconsistencies in the data rather than on specific detail.

In the last analysis, questions regarding the validity of GRE tests should be thought of and treated as recurring questions to which up-to-date answers should be sought frequently and locally. These first-year studies should be replicated, additional criteria might well be examined profitably, and studies involving longer-term criteria are important to the establishment of an informed basis for interpreting GRE scores in graduate school admissions contexts.

The findings of this GRE Cooperative Study will be of greatest value if perceived as first approximations in an iterative GRE validity-study process that is essentially open-ended.

---

*Tests of statistical significance have not been stressed in evaluating the validity coefficients obtained in these preliminary studies. A positive relationship between academic predictors and academic criteria is expected a priori, hence the null hypothesis (i.e., that no correlation exists between the predictor and criterion variables under consideration) is not deemed to be appropriate. Following completion of the individual institutional-departmental analyses and reports, summary distributions of obtained validity coefficients will be prepared in order to assess the range of observed values.

Description of the Samples, Data, and Procedures

The present study is based on data for 25 students in Geology/Geophysics and 22 Botany/Microbiology students in the University Graduate School. Both departments offer doctoral programs, and the sample studied includes both prospective master's- and prospective doctoral-degree candidates for whom first-year programs and evaluation procedures reportedly are comparable. Data were supplied for first-time enrolled, full-time, degree-seeking students who entered in fall 1974 and fall 1975, respectively.[*] The study focusses on the relationship between GRE scores and other measures of personal and background characteristics of students (potential predictors of perforamnce) and measures of their performance (criteria of "success") during the first year of graduate study.

Enumerated and described briefly below are the variables for which observations were reported and the number of students with observations on each variable:

Admissions variables	Geology/Geophy (n)	Botany/Microbiol (n)
GRE-Verbal [(GRE-V) 200-900]	24	10
GRE-Quantitative [(GRE-Q) 200-900]	24	10
GRE-Advanced Test (200-990)	Geology (23)	n.a.
Undergraduate GPA (UGPA) [4 = A, 3 = B, 2 = C, 1 = D, F = 0]	25	n.a.
Criterion (Performance) variables		
Graduate GPA, General (GGPA), Year 1 [scale as for UGPA]	25	22
Ad hoc Rating--potential for advanced study in the field. Average of faculty ratings on a four-point scale (see Table 3).	25	n.a.
Other variables		
Birth Year (last two digits)	25	n.a.
BA/BS Year (last two digits)		

---

[*] It is assumed that all such students were included on the data rosters provided.

Information regarding ethnicity was supplied for the Geology sample (which included only two non-Caucasian students, both of whom were classified as "Oriental or Asian-American") but not for the Botany sample. Twenty of the 25 Geology students and 17 of 21 Botany students for whom sex coding was provided were male.

All members of the entering cohorts being studied were classified as continuing in the department, in progress toward a degree, as of the beginning of the second year following admission, indicating no first-year attrition (assuming that all members of the entering cohort are represented in the respective samples).

For Geology sample data were provided on year of birth, year of bachelor's degree, and continuity/discontinuity with respect to major field and institution from undergraduate to graduate school.

- 22 of 25 students earned the bachelor's degree in 1974 (11) or 1975 (11); of the remaining three students, two were 1972 graduates and one graduated in 1968.
- In terms of age (year of birth) the sample was distributed over a somewhat wider range (approximately 20 to 29 years of age at time of entry into the department).
- Of the 25 students, five students were graduates of the University (continuity of institution) and 19 had an undergraduate major in Geology (continuity of field).

Limitations of the data

Apart from the small size of the samples under consideration, observations were not available for all individuals in each sample; GRE Aptitude Test scores were not available for 12 of the 22 individuals in the Botany/Microbiology sample. In small samples in which a predictor is not available for the entire entering group, evaluation of the validity of that predictor (and others) is doubly complicated. It cannot be assumed, for example, that the individuals for whom a predictor is available are "like" those for whom it is not available in terms of performance on other potential predictors. Individuals with "marginal" Undergraduate GPA, for example, may be required or may elect to supplement their admissions application with GRE Advanced or Aptitude Test scores; those with "very low" GRE-Verbal scores may have been admitted, in part, on the basis of "compensatorily high" GRE-Quantitative scores, etc.[*]

Even when all predictors are available for all individuals, there may be systematic effects on the observed validity coefficients as a consequence of the way in which the variables were employed in screening candidates.

The Graduate GPA, reflecting grades received for work completed during the first year of graduate study, is distributed over a very limited range. In Geology, for example, only one student received a GPA below 3.00 on a 4-point scale; only two of 22 Botany/Microbiology students earned less than a 3.00 GGPA (and GRE Aptitude scores are not available for one of these "low performing" students).

In view of the foregoing limitations, the findings reported herein clearly should be thought of as "first approximations" in a continuing validity assessment process.

Procedures

Due to the missing data pattern, descriptive statistics and validity coefficients were determined by using all observations available for a given variable or pair of variables. Thus, for example, the GRE-Aptitude scores for Geology are based on 24 cases, mean Advanced Geology score is based on only 23 cases, and the validity coefficients for these scores are based on 24 and 23 cases, respectively, having both the GFE scores and one or more criterion scores. In the Botany sample, GRE Aptitude validity coefficients were determined for the 10 students with GRE scores.

---

[*] See Robyn M. Dawes. Graduate admission variables and future success. Science, 1975, 187, 721-723, for an analysis of the attenuating effect on predictive validity of compensatory methods of screening applications for admission--i.e., selection using multiple assessment variables in such a way that if the selected individuals are low on any particular variable, they will tend to be compensatorily high on others.

Characteristics of the Samples

Table 3 shows measures of central tendency (the arithmetic average or mean) and variability (the standard deviation) for the distributions of predictor and performance (or criterion) variables in the respective samples.

- Both departmental samples are characterized by somewhat higher GRE-Q than GRE-V means; the average individual in these samples ranks slightly below the 60th percentile on GRE-Verbal, while the average GRE-Q scores rank somewhat lower than the 75th percentile nationally.

- Standard deviations of GRE-V scores are comparable for the two samples; however, with respect to quantitative ability, the Geology/Geophysics sample is more homogenous than the Botany/Microbiology sample, judging from the smaller standard deviation of GRE-Q scores in Geology. It is important to note, however, that GRE Aptitude scores are not available for more than half the Botany sample.

- With respect to scores on the GRE Advanced Test (Geology), the Geology sample is quite homogeneous (SD = 63) around a mean value that is about average for candidates who take this test nationally.

It was noted earlier than the Graduate GPA (first-year work) was distributed over a limited range--only one Geology and two Botany students received a GGPA of less than 3.00 on a four-point scale. The mean values of GGPA (3.47 and 3.45) reflect the dearth of "lower" grades in the respective departmental samples. Judging from standard deviations, the GGPA distribution for Botany is somewhat less homogeneous than that for Geology.

Variables on which observations are available for Geology but not Botany include the GRE Advanced Test, considered above, the Undergraduate GPA, Ad hoc faculty ratings, Birth Year and Year of Bachelor's Degree.

The Graduate GPA, of course, reflects more or less "routine" patterns of evaluation and grades in Geology appear to involve primarily "A's and B's," (i.e., 4's and 3's on a four-point scale). For the Geology sample, ratings of

Table 3

Measures of Central Tendency and Variability

Variable	Geology/Geophysics			Botany/Microbiology		
	N	Mean	Standard deviation	N	Mean	Standard deviation
GRE-Verbal	24	522	105	10	536	101
GRE-Quantitative	24	606	81	10	591	100
GRE-Advanced (Geology)	23	559	63	--	--	--
Undergraduate GPA (UGPA)	25	2.99	0.24	--	--	--
Graduate GPA (GGPA)	25	3.47	0.35	22	3.45	0.55
Ad hoc rating (average)*	25	2.49	0.51	--	--	--
Birth year	25	51.20	5.04	--	--	--
BA/BS year	25	74.04	1.48	--	--	--

*Rated in terms of potential for advanced study on a four-point scale:
4 = Outstanding performer, definitely qualified for doctoral study;
3 = Definitely master's caliber; probably capable of acceptable doctoral study;
2 = Adequate to adequate plus at master's level; would not encourage doctoral;
1 = Unacceptable or marginally acceptable for graduate study at master's level.

Table 4

Correlation of Predictors with Criterion Variables

Variable	Geology/Geophysics			Botany/Microbiology	
	N	GGPA	Ad hoc rating	N	GGPA
GRE-Verbal	24	.13	.43	10	.18
GRE-Quantitative	24	-.03	.26	10	.42
GRE-Advanced (Geology)	23	.14	.40	--	--
(V + Q)/2 or GRE Aptitude Average	24	.06	.37	10	.34
Undergraduate GPA	25	.27	-.34	--	--
[Graduate GPA]	[25]	[1.00]	[.33]	--	--
Birth year	25	-.25	.20	--	--
BA/BS year	25	-.33	-.07	--	--

potential for advanced study were provided on a four-point scale.* Given the new frame of reference reflected in the ad hoc ratings (see note to Table 3), faculty judgments of student performance were distributed over a wider range (standard deviation of .51 as compared to .24 for GGPA).

Correlation of Predictors with Criterion Variables

Table 4 shows correlation coefficients indicating the relationship of the respective admissions variables to Graduate GPA and Ad hoc ratings, respectively, in the Geology sample. Comparable coefficients for GRE-V and GRE-Q, involving only 10 cases with these scores, are shown for Botany/Microbiology, along with a coefficient for a variable called "GRE Aptitude available" (coded "1" if scores were present). In both samples, a coefficient is reported for GRE(V + Q)/2 (the average of an individual's V and Q scores, the operational equivalent of the GRE total score more frequently encountered in practice).

Findings for Botany

For 10 students with GRE Aptitude scores, GRE-Quantitative and GRE-Verbal are both positively related to Graduate GPA, with GRE-Q somewhat more closely associated ($r = .42$) than GRE-V ($r = .18$). The positive coefficient ($r = .21$) for "GRE Aptitude Available" indicates a tendency for Graduate GPA to be higher for individuals with GRE Aptitude scores than for individuals without GRE Aptitude scores. The nature of the relationship is shown below:

	Graduate GPA		
GRE Availability	Below 3.50	3.50 - 3.74	3.75 plus
GRE scores available (10)	2	2	6
GRE scores not available (12)	6	5	1
Total	8	7	7

---

*No description was provided of the procedures employed in developing ad hoc rating. In the absence of explicit information about procedures, it is assumed that the ratings were made in accordance with the standard schedule for rating students in terms of potential for advanced study in a field.

Findings for Geology

The validity coefficients for GRE predictors in Geology suggest (a) that Faculty Ratings tend to be more closely associated with GRE scores than is the General GPA, and (b) that GRE Advanced Geology Test and GRE-Verbal scores may tend to be more closely associated with performance criteria among enrolled students than GRE-Quantitative.

- Coefficients for GRE-Verbal are positive with respect to both GGPA and Ad hoc ratings, but somewhat larger when Ad hoc ratings (r = .43) constitute the criterion than when GGPA is the criterion (r = .13). GRE Advanced Geology has a comparable pattern of validities (r = .40 vs. Ad hoc ratings; r = .14 vs. GGPA). A small negative coefficient was obtained for GRE-Q vs. GGPA (r = - .03); when Ad hoc ratings constituted the criterion, the coefficient for GRE-Q was positive (r = .26).
- The Undergraduate GPA is positively associated with Graduate GPA but negatively associated with Ad hoc ratings in this sample. This finding must be considered anomalous. The expectation, a priori, is that past academic performance will tend to be positively associated with future academic performance and that different measures of performance will tend to be positively associated.

The two criterion (performance) measures are positively associated although the relationship is relatively low (r = .33, Table 4). Lengthy speculation about "reasons" for findings of this type is not warranted here. However, it is useful to note that in small samples one or two "aberrational" data sets can have considerable influence on outcomes. In the present Geology sample, for example, two individuals had unusually high Undergraduate GPA (3.65 and 3.61, compared to the mean value of 2.99, standard deviation of .24; only two other UGPA greater than 3.00)--both of these individuals were among the four lowest ranking students in terms of Ad hoc ratings and, thus, contributed substantial negative covariance in the UGPA/Ad hoc rating analysis.*

---

* By examining the original data rosters, the Department can identify the two cases involved. It would be of some interest to look into the records of the students involved in an effort to identify circumstances that might help to "explain" the exceptionally high undergraduate record and the "very low" Ad hoc ratings. The possibility of clerical error should be considered in such an examination. Perhaps the undergraduate institutions had somewhat "lower standards," etc.

No a priori set of expectations can be adduced to guide evaluations of the pattern of coefficients for Birth Year (inversely related to age) and Year of Bachelor's Degree. They are shown only as descriptive of tendencies within the present sample. In this sense, with respect to General Graduate GPA, the negative coefficient ($r = -.25$) for Birth Year indicates a tendency for younger students to earn lower GPA, and the positive coefficient ($r = .20$) with respect to Ad hoc ratings indicates the opposite tendency. In the case of Year of Bachelor's Degree, with respect to both criterion variables there is a tendency in the sample for more recent graduates to receive lower criterion scores. However, as suggested above, there is no a priori rationale for evaluating these findings and they should not be generalized.

Validity of Composites

In view of the small size of these samples and the anomalous patterns of coefficients there was no application of formal multivariate procedures for establishing the relationship of two or more predictors to the respective criterion variables. However, it is of interest to call attention to the relationship of GRE(V + Q)/2 to the criterion variables (cf., pp. 7-8, for discussion of weighting). It may be noted (Table 4) that this simple average of GRE-Verbal and GRE-Quantitative yields lower validity coefficients with criterion variables than GRE-V or GRE-Q, whichever is higher. When GRE-V and GRE-Q, as in the present samples, have quite different individual validities with respect to a criterion variable, the GRE Total (or GRE Average) will tend to have less validity than one of its components--the sum thus has less validity than one of its parts. The point to be made here is that composites of GRE scores that have apparent "face validity" (i.e., appear to be logical) may not in fact be appropriate. Empirical validation is needed for all GRE composites.

In the present exploratory study, questions of relative weighting for GRE-V, GRE-Q, (and UGPA in the case of Geology) cannot be addressed directly. It is, however, of interest to call attention to the possible impact of restriction of range within departmental samples on the patterns of validity coefficients for predictors. In the present situation, for example, GRE-Q has lower correlation with criterion variables than GRE in the Geology sample and higher correlation than GRE-V in the Botany sample. Questions of sampling fluctuation aside, it is relevant to note that the standard deviation for GRE-Q is considerably smaller than that for GRE-V in the Geology sample whereas comparable standard deviations obtain for these two predictors in the Botany sample. The Geology sample is more restricted with respect to spread of GRE-Q than with respect to GRE-Verbal scores, a condition which tends to be associated with somewhat lower validities for the variable characterized by restriction.

Conclusions and Recommendations

As indicated at the outset, this study should be thought of as yielding first approximations in an iterative GRE-Validation process that is essentially open-ended. The findings suggest that GRE scores contain information of value when considered along with other information used in screening applicants for admission. An anomalously negative validity for Undergraduate GPA with respect to one of the criterion variables should not be thought of as providing a basis for generalization--such a coefficient in a small sample may be influenced by one or two aberrational data-sets. Attention has been directed to two such data-sets in the Geology sample.

The need for replication of validity studies is clearly indicated. Possible directions for further examination of the validity of admissions variables (GRE and others) are suggested below:

- The highly restricted nature of the Graduate GPA distributions involved has been noted. The fact that grades tend to be distributed primarily over a one-point range in a four-point grade scale tends to limit the utility of this variable as a criterion. Faculty ratings appear to be a more predictable criterion than Graduate GPA. Results of this study suggest the importance of further exploration of the utility of systematic faculty ratings of students in terms other than those specified by routine grades-in-course.

- A variety of rating procedures may be thought of as being "administratively viable" and such procedures would permit assessment of student performance (if only for purposes of research and evaluation) outside the traditional GPA- system that is conceptually more applicable to undergraduate than to graduate settings.
- Replication of one-year validity studies such as those reported herein would be facilitated if rosters of first-time enrolled students enrolling each year were developed by each department. GRE scores, UGPA, and other potentially important admissions data for each such student could be recorded. These rosters could be updated with General Graduate GPA and other performance data, including systematic faculty ratings, at the end of the first year of graduate study (and subsequent years of study, as well).

Kenneth M. Wilson
June 1977
Princeton, NJ

Appendix C-2

GRE COOPERATIVE VALIDITY STUDIES PROJECT

Educational Testing Service
Princeton, NJ 08541

To: Participating Graduate Schools
Date: March 1978
From: Kenneth M. Wilson

Subject: Tabular Summary of Selected Validity Study Findings

Attached is a tabular summary of selected findings of studies that have been completed as part of the GRE Cooperative Validity Studies Project, sponsored by the Graduate Record Examinations Board. The Cooperative Studies have been concerned with the relationship of GRE Test scores and/or Undergraduate Grade Point Average (UGPA), as available, to First-Year Graduate Grade Point Average, as well as other criteria of performance during the first year of graduate study. To date, validity study reports have been prepared for 34 graduate schools; for several other schools data are currently being analyzed and it is expected that a few more schools will provide data (see list). The data in this summary reflect selected findings of the 34 completed studies. The summary is intended to provide general normative perspective to facilitate assessment and evaluation of trends and patterns across graduate schools and departments.

The "common criterion" in the Studies that have been completed has been the first-year Graduate GPA (GPA General). Other criteria such as ratings or examination scores are very unevenly available. Table A summarizes information regarding the number of samples for which data have been analyzed by discipline or department. The number of samples varies considerably across disciplines. This table also shows weighted mean values of validity coefficients for GRE-Verbal (GRE-V), GRE Quantitative (GRE-Q), GRE-Advanced (GRE-Adv), and Undergraduate GPA (UGPA), as available.

The weighted mean values reflect degree of within-group covariation between relative standing on the respective predictors and relative standing on the Graduate GPA criterion in pooled samples of all individuals with predictor and criterion scores. The coefficients indicate what the predictor-criterion relationship would be in such pooled samples after all predictor and criterion scores had first been standardized within each of the samples involved--ie., standardized within each departmental sample and then pooled for analysis.

In one or two instances a Critical Graduate GPA (e.g., grades in required or common sequences only) or an average of two or more criterion variables is involved, rather than the Graduate GPA General. Some clustering of field/departments has been introduced, as indicated in notes to Table A. The data in Table A are of interest from several points of view:

1. Numbers in parentheses following the number of departmental samples involved indicate the number of samples with Ns for validity coefficients that were equal to or greater than the number suggested as the minimum target for the Cooperative Studies, namely, N = 25, which by design was to be reached by combining data for two successive entering cohorts of first-time enrolled, full-time, degree-seeking students. In the aggregate, 86 of 137 Aptitude coefficients (about 63 percent), 56 of 81 UGPA coefficients (about 69 percent), and 27 of 69 Advanced Test coefficients (about 39 percent) were based on 25 or more cases. The remainder were based on fewer than 25 cases.

2. The pattern of coefficients suggests the potential importance of the predictor that frequently was "least available", namely, GRE Advanced Tests. Advanced Test scores tended to be missing for a number of individuals in the samples under consideration. The influence of availability versus unavailability of the GRE Advanced Test score (or scores on other predictors) on the observed patterns of validity coefficients cannot be estimated. In a number of instances, more than one Advanced Test (field) was represented in a data-sample e.g., a chemistry sample may have included not only students with Chemistry scores but also one or more with Mathematics, Physics or Engineering. Despite missing data limitations and occasional Advanced Test-field heterogeneity, the weighted mean coefficients for the Advanced Tests (which reflect aptitude, motivation, and substantive achievement) suggest their potential importance.

3. Generally speaking, the average values of the coefficients in Table A clearly are consistent with the working proposition that a positive association exists between measures of developed ability and achievement (such as GRE Tests and UGPA) and measures of first year performance in graduate study, such as the Graduate GPA general.

In evaluating the magnitudes of the coefficients, it is important to recognize that the Graduate GPA criterion used frequently (typically) had a severely restricted range, and was sometimes simply dichotomous and heavily weighted in the direction of "passing" marks.* In the circumstances, the size of the pooled within-group average values of validity coefficients undoubtedly are lower than would be the case if differences in student performance were more rigorously and reliably reflected in the "routine" grading process.

---

*When dichotomous GPA criteria were used, point biserial coefficients were computed. Since the underlying criterion variable involved was actually continuous, the point biserial underestimates the relationship of the continuously distributed predictor to the underlying continuous criterion variable.

Summarization by Field and School

The data in Table A provide evidence regarding the typical levels and patterns of coefficients by field. Data summarized in Tables 1 through 11 show, for designated fields or groups of fields/departments (a) means and standard deviations of predictor and criterion scores in each sample, (b) information regarding sample size, and (c) the observed validity coefficients.

Validity coefficients are not reported if based on less than N = 10. In such cases, the sign of the observed coefficient is reported to indicate the direction of the relationship in the sample. It is important to note, however, that (a) the typical level and pattern of coefficients in these very small samples followed closely that for all samples as reported in Table A, and (b) that the weighted coefficients derived from these extremely small samples were included in developing the weighted mean values reported in Table A.

Most Graduate GPA and Undergraduate GPA scales were A = 4, B = 3, C = 2, etc.; occasionally the GPAs reported by schools were on other scales. For purposes of the summary presentation in Tables 1 through 11, means and standard deviations based on atypical scales have been adjusted to make them roughly comparable.

In some instances, multiple correlation coefficients based on missing data correlation procedures are shown (in parentheses under the UGPA Column). Unless otherwise indicated these coefficients are based on all the predictors for which validity coefficients are reported. Multiples are not routinely reported due to sample-size, missing data and other related considerations. The purpose in reporting multiple correlations in these small samples is primarily to provide some perspective on the potential value of combining two or more predictors.

Several features of the data in Tables 1 through 11 are noteworthy, including the following:

1. Despite the limitations of missing data procedures, the multiple correlation coefficients suggest that the test variables and the undergraduate grade point record tend to provide some uniquely important information about student performance-potential.

2. In examining the tables it will be seen that "useful" levels of within-group validity are to be found in data for samples differing considerably in level of scores on the GRE predictors.

3. The interdepartmental data provide useful normative perspective regarding the range and patterning of validity coefficients (including occasional negative coefficients). The potential value of the GRE-Advanced Tests, suggested by the average values reported in Table A, is also suggested in the individual departmental analyses where coefficients for the other available predictors may be seen.

It is important to keep in mind that the coefficients in Tables 1 through 11 have relatively large sampling errors due to the small size of the respective samples. Conclusions regarding the relative validity in particular samples of V, Q, and Advanced Tests, and UGPA call for the accumulation of a substantial body of empirical evidence derivable only through replication of validity studies.

Graduate Schools Participating in Cooperative Studies

School
Air Force Institute of Technology
Auburn University
Baylor University
Bradley University
Brown University
California State University at Fullerton
Florida Technological University
Fort Hays Kansas State College
Harvard University
Hofstra University
Indiana University at Bloomington
Louisiana State University
Loyola University at Chicago
The Ohio State University
Old Dominion University
Oregon State University
Princeton University
Stanford University
State University of New York at Stony Brook
University of Arizona at Tucson
University of California at Berkeley
University of Colorado at Boulder
University of Hawaii at Honolulu
University of Illinois
University of Kentucky
University of Massachusetts
University of Miami
University of Michigan
University of Missouri at Rolla
University of Montana
University of New Orleans
University of North Carolina at Chapel Hill
University of Notre Dame
University of Oklahoma
University of Texas at Arlington
University of Virginia
University of Wisconsin
Virginia State College
Washington State University at Pullman
Wayne State University

Table A

Summary of Data Submitted and Observed Validity Patterns, By Field: Grad GPA Criterion

Field/ Department	Number of Samples						Weighted Mean Coefficients				
	GRE-V [Aptitude]	GRE-Q	GRE-Adv		UGPA		GRE-V	GRE-Q	GRE-Adv	UGPA	
Biosciences[a]	22	(12)*	13	(2)	14	(5)	.19	.25	.37	.24	
Chemistry	12	(6)	7	(5)	8	(7)	.06	.25	.39	.31	
Engineering[b]	10	(4)	4	(0)	5	(2)	.28	.30	.28	.20	
Mathematics[c]	6	(3)	2	(0)	2	(0)	.32	.23	.35	.30	
Physics	5	(3)	4	(3)	2	(2)	.05	.16	.19	.29	
Geol, Geophysics	5	(1)	4	(1)	1	(1)	.05	.06	.11	.37	**
Economics	6	(4)	3	(3)	3	(3)	.09	.34	.45	.27	
Anthropolgy	3	(2)	-	-	1	(1)	.26	.21	-	.06	**
Education[d]	7	(6)	2	(2)	5	(5)	.18	.12	.54	.24	
English	6	(3)	5	(0)	4	(2)	.41	.24	.48	.22	
History	10	(10)	7	(3)	8	(8)	.31	.26	.21	.30	
Pol Science[e]	4	(4)	2	(1)	3	(3)	.43	.34	.49	.18	
Psychology	12	(10)	7	(5)	7	(4)	.24	.26	.37	.22	
Sociology[f]	7	(5)	3	(1)	5	(4)	.43	.30	.54	.55	
Library Sci	3	(3)	-	-	3	(3)	.32	.52	-	.33	
Fine Arts[g]	6	(6)	-	-	5	(5)	.33	.26	-	.31	
Music	3	(3)	2	(1)	1	(1)	.24	.11	.21	.23	**
Philosophy	5	(0)	2	(0)	2	(0)	.25	.14	.23	.56	
Languages[h]	5	(1)	2	(0)	2	(0)	.31	.20	.45	.28	

NOTE: The validity coefficients shown are weighted averages of obtained coefficients. Patterns of medians are similar.

*Ns in parentheses indicate the number of samples for which N=25 or greater, based on data for two years--1974-75 and 1975-76 in almost every instance.

**Coefficient based on one sample only.

[a] Includes Oceanography, Marine Environmental Science, Allied Health Science

[b] Includes Engineering and Facilities Management

[c] Includes Computer Science, Applied Math and Stat

[d] Includes Vocational and Adult Education, Educ Administration

[e] Includes Public Administration

[f] Includes Social Work, Urban Planning, Public Policy Studies

[g] Includes Speech and Theater, Drama & Communication, Speech & Comm, and Journalism

[h] Includes two Hispanic, one Germanic, one French, and one undifferentiated Foreign Languages & Literatures

Table 1 Page 1 of 2 pages

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

BIOLOGICAL SCIENCES

	School (coded) (N / coefficient)	Predictor mean & s.d.				Validity coefficient				Grad GPA mean & s.d.
		GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	
						Zoology-Fisheries				
M[a]	School 007	524	580	—	2.99	.21	.26	—	.24	3.59
	(80,80,—,80)	92	88		0.39				(R = .43)	0.34
						Biology				
M&D	School 009	(62)	(70)	(59)	3.34	.27	.47	.54	.32	3.60
	(17,17,15,17)	25	17	23	0.35				(R = .66)	0.31
			Percentiles			Biology				
M&D	School 046	620	665	754	—	.06	.02	.01	—	3.81
	(21,21,18,—)	99	141	109						0.21
						Botany				
M&D	School 080	598	591	660	—	.32	.43	.20	—	3.70
	(15,15,14,—)	64	89	87						0.30
						Zoology				
M&D	School 080	620	652	—	3.31	-.03	.02	—	.15	3.20
	(19,19,—,19)	92	59		0.29					0.60
						Medical Center				
M	School 087	565	630	615	—	-.08	.11	.41	—	3.05
	(26,26,16,—)	106	85	88						0.66
						Biology				
M&D	School 097	551	653	686	3.36	.05	-.07	.26	.16	3.41
	(26,26,22,32)	76	78	66	0.33					0.36
						Biology				
M	School 122	495	566	600	2.91	.48	.32	.61	.35	2.95
	(25,25,25,25)	105	109	104	0.36				(R = .62)	1.12
						Zoology				
M&D	School 123	540	591	662	3.15	.24	.56	.56	-.07	3.42
	(29,29,20,28)	118	93	98	0.38					0.61
						Forestry				
M+	School 123	473	555	595	3.03	.38	.18	.41	.31	3.40
	(43,43,11,43)	106	94	93	0.32					0.45
						Biology				
M&D	School 145	617	674	--	3.41	.29	.37	--	.23	3.56
	(31,31,--,37)	96	79		0.33				(R = .44)	0.53
						Natural Resources				
M&D	School 145	587	612	—	3.28	.33	.31	—	.21	3.42
	(38,38,--,38)	82	98		0.43				(R = .38)	0.53
						Forestry				
M	School 149	567	630	--	—	.03	.36	--	--	3.60
	(33,33,--,--)	77	87							0.31

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

[a] Indicates degree orientation of students in sample: M = master's, M&D = both master's and doctoral candidates included; M+ = largely master's, with several post-master's doctoral students included.

Table 1 (con't) Page 2 of 2 pages

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

BIOLOGICAL SCIENCES

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
						Botany				
?[a]	School 169	592	609	714	3.28	.29	.29	(+)*	.40	3.75
	(13,13, 8,13)	102	78	90	0.45					0.21
						Microbiology				
?	School 169	566	600	577	3.13	(+)*	(+)*	(+)*		3.61
	(6,6,3,6)	81	75	54	0.38					0.28
						Zoology				
M&D	School 169	609	663	733	3.17	-.06	.22	.21	.45	3.59
	(15,15,15,14)	62	57	67	0.34					0.27
						Biology				
M&D	School 147	510	588	628	3.19	.10	.35	.37	.34	2.88
	(43,43,42,52)	86	90	83	0.34				(R = .46)	0.54
						Marine Environmental Science				
M	School 204	572	656	—	—	-.07	.21	—	—	3.42
	(21,21,—,—)	91	71							0.56
						Zoology				
M&D	School 231	564	621	—	3.28	.60	.72	—	-.13	3.65
	(19,19,—,19)	96	94	97	0.35					0.40
						Allied Health Science				
M	School 231	529	567	—	—	.07	.06	—	—	3.00
	(43,43,—,—)	86	105							0.73
						Botany/Microbiology				
M&D	School 248	536	591	—	—	.18	.42	—	—	3.45
	(10,10,—,—)	101	100							0.55
						Oceanography				
M&D	School 251	589	677	—	3.42	.15	.07	—	.31	3.51
	(34,34,—,34)	105	82		0.28					0.29
						Biology				
M	School 293	556	594	679	—	.80	-.03	.13	—	3.66
	(11,11,10,—)	94	60	73						0.32
						Speech and Hearing Science				
M	School 009	530	551	—	3.31	.28	.34	—	.44	3.37
	(22,22,—,22)	62	73		0.37				(R = .56)	0.41
						Speech & Hearing - GPA Gen. & GPA Crit.				
M	School 095	468	451	—	3.14	.15	.05	—	-.03	3.73
	(20,20,—,20)	48	91		0.28					0.23

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

[a] Indicates degree orientation of students in sample: ? = not indicated by department, M&D = both prospective master's and doctoral students, M = master's only.

Table 2

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

CHEMISTRY

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N/coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
M[a]	School 293	433	649	—	—	.28	.04	—	—	3.04
	(10,10,—,—)	111	118							0.57
?	School 169	579	661	565	3.16	.40	.15	(+)*	.03	3.23
	(14,14, 8,14)	105	73	79	0.43					0.46
M&D	School 118									
	Foreign	320	687	—	—	.09	.59	—	—	3.19
	(19,19,—,—)	80	79							0.60
	Nonforeign	514	626	—	—	.21	.43	—	—	3.00
	(21,21,—,—)	97	80							0.55
M&D	School 204	497	648	663	—	-.18	.32	.55	—	3.18
	(29,29,17,—)	154	104							0.50
M&D	School 251	--	--	--	3.42	--	--	--	.44	3.28
	(--,-----,39)				0.35					0.54
M&D	School 221	616	692	667	3.38	-.21	.17	.28	.37	3.12
	(52,52,43,49)	94	75	71	0.38					0.31
M&D	School 145	601	706	—	3.28	.41	.46	—	.36	3.02
	(29,29,—,48)	94	80		0.29				(R = .51)	0.69
M&D	School 046	533	657	671	3.25	.00	.37	.41	.11	3.48
	(15,15,14,27)	114	82	111	0.35				(R = .54)	0.56
M&D	School 009	603	698	669	3.43	.22	.52	(+)*	.21	3.62
	(12,12, 9,26)	71	66	85	0.33					0.47
?	School 231	525	654	680	3.27	-.21	.50	.59	.58	3.29
	(43,43,41,32)	99	86	92	0.38					0.34
M&D	School 097	581	705	700	3.54	.19	.34	.31	.29	3.49
	(94,94,87,133)	96	83	77	0.29				(R = .47)	0.41

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

[a] Indicates degree orientation of students in sample: M = master's, ? = not designated by department, M&D = both prospective master's and doctoral students.

Table 3

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

ENGINEERING AND RELATED FIELDS

	School (coded) (N/coefficient)	Predictor mean & s.d.				Validity coefficient				Grad GPA mean & s.d.
		GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	
						Engineering				
M[a]	School 069	497	633	558	3.26	.25	.10	(-)*	-.13	3.55
	(32,32,08,12)	87	123	72	0.25					0.31
						Chemical Engineering (for.)				
M&D	School 118	352	640	—	—	(+)*	(+)*	—	—	3.40
	(5,5,—,—)	35	82							0.11
						Chemical Engineering (nonfor.)				
M&D	School 118	514	694	—	—	.34	.46	—	—	3.49
	(14,14,—,—)	75	56							0.36
						Engineering Management (for.)				
M	School 167	306	601	455	3.12	.26	.48	.74	(+)*	3.60
	(11,11,11,4)	74	106	68	0.40					0.20
						Engineering Management (nonfor.)				
M	School 167	485	675	513	2.94	.12	-.30	-.03	.41	3.62
	(15,15,15,15,)	68	64	97	0.45					0.26
						Electrical Engineering				
M&D	School 204	458	681	709	—	.24	.42	(+)*	—	3.41
	(32,32,9,—)	120	102	100						0.27
						Electrical Engineering				
M	School 246	556	679	—	2.92	.48	.34	—	.61	3.51
	(30,30,—,30)	107	65		0.39					0.38
						Facilities Management				
M	School 246	507	561	—	2.56	.26	.38	—	-.25	3.72
	(30,30,—,30)	64	78		0.35					0.16
						Civil Engineering (for.)				
M&D	School 293	285	652	—	—	.11	.30	—	—	3.46
	(19,19,—,—)	66	66							0.41
						Civil Engineering (nonfor.)				
M&D	School 293	433	649	—	—	.30	.47	—	—	3.57
	(14,14,—,—)	101	67							0.52

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

[a] Indicates degree orientation of students in sample: M = master's, M&D = both prospective master's and doctoral students.

Table 4

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

MATHEMATICS, COMPUTER SCIENCE & ECONOMICS

	School (coded) (N / coefficient)	Predictor mean & s.d GRE-V	Predictor mean & s.d GRE-Q	Predictor mean & s.d GRE-Adv	Predictor mean & s.d UGPA	Validity coefficient GRE-V	Validity coefficient GRE-Q	Validity coefficient GRE-Adv	Validity coefficient UPGA	Grad GPA mean & s.d.
						Mathematics				
D[a]	School 101	574	744	741	3.71	.36	-.01	.58	.32	3.03
	(13,13,12,14)	74	25	90	0.30					0.41
						Computer Science				
M&D	School 132	610	704	(675)*	—	.25	.08	.22	—	3.64
	(43,43,22,—)	98	78	119						0.55
						Chem/Mathematics				
M	School 169	599	687	—	3.25	.55	.40	—	.29	3.35
	(18,18,—,18)	108	83		0.45					0.47
						Mathematics				
M&D	School 204	538	703	—	—	.24	.40	—	—	3.37
	(12,12,—,—)	115	86							0.53
						Applied Math				
M&D	School 204	488	661	—	—	.27	.27	—	—	3.32
	(40,40,—,—)	125	102							0.49
						Computer Science				
M&D	School 204	547	691	—	—	.34	.31	—	—	3.79
	(28,28,—,—)	137	94							0.41
						Economics				
D	School 033	680	732	790	—	-.14	.01	.49	—	3.08
	(37,37,26,—)	97	102	93						0.59
						Economics				
M&D	School 097 **	56	75	65	3.53	.05	.29	.53	.16	3.60
	(43,43,43,[illegible])	26	21	23	0.35					0.36
						Economics				
M&D	School 118	508	610	—	—	.06	.27	—	—	3.46
	(20,20,—,—)	99	76							0.26
						Economics				
M&D	School 145	597	700	—	3.54	.17	.37	—	.02	3.15
	(36,36,—,37)	74	65		0.31				(R = .38)	0.40
						Economics				
M&D	School 204	473	585	—	—	.37	.56	—	—	2.91
	(19,19,—,—)	136	121							1.00
						Economics				
M&D	School 320	587	672	645	3.16	.14	.55	.33	.53	3.46
	(49,49,41,50)	102	85	85	0.54				(R = .71)	0.36

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Eight different Advanced Test fields, largely Mathematics; no Computer Science Scores.

**Percentile ranks

[a]Indicates degree orientation of students in sample: D = prospective doctoral students only, M&D = both prospective master's and doctoral students, M = prospective master's only.

Table 5

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

PHYSICS, GEOLOGY, GEOPHYSICS

	School (coded) (N / coefficient)	Predictor mean & s.d. GRE-V	GRE-Q	GRE-Adv	UGPA	Validity coefficient GRE-V	GRE-Q	GRE-Adv	UPGA	Grad GPA mean & s.d.
						Physics				
D[a]	School 035	613	746	815	—	.06	-.04	.10	—	3.36
	(27,27,28,—)	108	67	122						0.48
						Geology				
D	School 035	634	746	685	—	-.01	-.06	.06	—	3.72
	(16,16,15,—)	87	58	89						0.18
						Geophysics				
D	School 035	615	743	728	—	.01	.32	(-)*	—	3.83
	(10,10, 5,—)	110	44	58						0.39
						Physics				
M&D	School 046	588	720	744	3.52	.02	.16	.31	.23	3.57
	(41,41,40,43)	99	69	97	0.28				(R = .40)	0.40
						Geology				
M&D	School 046	575	642	660	3.21	.09	.14	.18	.37	3.51
	(39,39,32,39)	105	76	82	0.36				(R = .37)	0.26
						Physics				
M&D	School 097	617	749	743	3.60	-.01	.17	.11	.31	3.44
	(98,98,99,108)	91	53	99	0.31					0.40
						Physics				
D	School 101	556	696	579	—	(+)*	(+)	(+)*	—	3.26
	(9,9,9,—)	111	79	97						0.57
						Geology/Geophysics				
M&D	School 248	522	606	559	2.99	13	.03	.14	.27	3.47
	(24,24,23, 24)	105	81	63	0.24					0.35
						Geology				
M	School 293	472	604	—	—	(-)*	(-)*	—	—	3.80
	(9,9,—,-)	62	71							0.31
						Physics				
M	School 293	579	715	—	—	(+)*	(+)*	—	—	3.57
	(8,8,—,—)	138	37							0.37

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

---

[a] Indicates degree orientation of students in sample: D = prospective doctoral candidates, M&D = both prospective master's and doctoral students, M = master's only.

Table 6

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

ENGLISH, SPEECH, COMMUNICATION, THEATER

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
D[a]	School 035 (21,21,19,—)	699 83	586 112	624 97	—	English .59	.15	.81	—	3.76 0.83
M/D	School 123 (24,24,17,23)	581 110	508 98	559 79	3.28 0.50	English .30	.21	.29	.26 (R = .38)	3.65 0.35
M&D	School 132 (19,19,18,23)	648 81	516 109	633 81	3.49 0.26	English .22	-.04	.40	.25 (R = .51)	3.63 0.29
M&D	School 145 (46,46,—,46)	663 66	548 101	—	3.60 0.32	English .23	.18	—	.08 (R = .38)	3.62 0.34
M&D	School 204 (28,28,20,—)	590 129	494 130	581 99	—	English .72	.45	.62	—	3.68 0.29
M&D	School 221 (52,52,48,52)	652 86	572 102	591 77	3.59 0.27	English .44	.34	.40	.29 (R = .59)	3.07 0.39
M	School 007 (40,40,—,40)	492 88	473 102	—	2.11 0.41	Speech and Communication .41	.42	—	.31 (R = .43)*	3.54 0.49
M	School 122 (31,30,—32)	477 117	458 111	—	2.91 0.36	Drama and Communication .27	.23	—	.30 (R = .36)	3.62 0.36
M+	School 123 (36,36,—,36)	522 90	477 99	—	3.07 0.31	Speech and Theater .34	.23	—	.27 (R = . 36)	3.42 0.49
M	School 123 (38,38,—,—)	476 118	433 78	—	—	Fine Arts .50	.13	—	—	3.56 0.41
M&D	School 145 (39,39,—,39)	537 97	509 108	—	3.50 0.26	Speech and Theater .35	.13	—	.30 (R = .43)	3.42 0.55
M	School 231 (60,60,—,72)	564 109	490 114	—	3.08 0.39	Journalism .21	.34	—	.35 (R = .43)	3,57 0.29
M&D	School 231 (44.44.33.--)	645 94	529 99	592 87		English .22	.15	.17		3.68

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*This coefficient is for a composite of predictors used in admissions.

[a] Indicates degree orientation of students in sample: D = prospective doctoral students only, M/D = master's plus some post-master's doctoral students, M = master's only, M+ = largely master's plus several doctoral students with master's degree.

Table 7

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

HISTORY, AREA STUDIES, ANTHROPOLOGY

	School (coded) (N / coefficient)	Predictor mean & s.d.				Validity coefficient				Grad GPA
		GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
D[a]	School 035	698	603	595	—	History .27	.39	.39	—	3.56
	(36,36,17,—)	78	98	54					(R= .44)	0.55
M&D	School 046	615	535	563	—	History .10	.20	.05	—	3.81
	(31,31,13,—)	87	98	60					(R= .21)	0.26
D	School 080	601	541	—	3.39	History .16	-.14	—	.02	3.51
	(36,36,—,38)	113	107		0.42					0.35
M	School 097	612	538	563	3.56	History .06	.26	-.10	.40	3.47
	(30,30,26,32)	87	94	76	0.34					0.51
M	School 103	594	552	557	3.42	History .38	.45	.15	.63	3.64
	(34,34,26,26)	91	121	55	0.38					0.31
M+	School 123	589	520	534	3.14	History .58	.36	.72	.38	3.47
	(27,27,14,27)	87	111	81	0.51				(R= .69)*	0.51
M&D	School 145	647	546	—	3.55	History .20	.20	—	.20	3.42
	(48,48,—,118)	99	126		0.35				(R= .29)	0.43
M	School 145	651	574	—	3.32	Asian Studies .66	.35	—	.12	3.33
	(37,37,—,55)	78	103		0.52				(R= .68)	0.46
M&D	School 221	649	602	594	3.62	History .29	.22	.36	.53	3.26
	(43,43,43,32)	80	107	73	0.28				(R= .62)	0.44
M+	School 231	603	509	570	3.34	History .42	.45	.02	.43	3.46
	(26,26,22,26)	109	121	79	0.50				(R= .56)	0.57
M&D	School 009	642	582	—	—	Anthropology .07	.11	—	—	3.62
	(37,37,—,—)	90	85							0.37
M&D	School 145	649	580	—	3.62	Anthropology .41	.30	—	.06	3.58
	(39,39,—,47)	118	109		0.30				(R= .46)	0.40
M&D	School 204	502	483	—	—	Anthropology .31	.20	—	—	3.45
	(19,19,—,—)	143	118							0.37

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on V, Q, and UGPA, only.

[a] Indicates degree orientation of students in the sample: D = prospective doctoral students, M&D = both prospective master's and doctoral students, M = master's students, M+ = largely master's plus several doctoral students with a master's degree.

Table 8

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

POLITICAL SCI, PUBLIC ADMINISTRATION, SOCIOL, URBAN PLAN, SOCIAL WORK

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
						Political Science				
M&D[a]	School 145	612	604	—	3.39	.44	.46	—	.30	3.44
	(30,30,—,39)	117	147		0.44				(R = .49)	0.42
						Public Administration				
M	School 221	527	505	461	3.16	.32	.24	.50	.04	3.10
	(40,40,14,39)	103	122	78	0.47				(R = .55)	0.37
						Political Science				
M&D	School 231	582	576	—	—	.33	.26	—	—	3.49
	(32,32,—,—)	114	136							0.38
						Govt. and Foreign Affairs				
M&D	School 320	611	569	551	3.44	.53	.38	.49	.18	3.59
	(79,79,61,69)	86	112	83	0.39				(R = .59)	0.32
						Sociology				
M&D	School 046	555	509	550	3.01	.41	.38	.51	.64	3.64
	(27,27,26,27)	132	128	97	0.54				(R = .76)	0.29
						Sociology				
M&D	School 097	600	551	584	—	(+)*	(+)*	(+)*	—	3.74
	(9,9,9,—)	69	89	85						0.20
						Sociology				
M	School 122	497	466	500	3.00	(+)*	(+)*	(+)*	(+)*	3.38
	(8,7,8,8)	128	67	77	0.31					0.48
						Public Policy				
M	School 145	597	625	—	3.41	.50	.69	—	.54	3.23
	(43,43,—,48)	96	125		0.41				(R = .80)	0.49
						Sociology				
M&D	School 145	601	606	—	3.42	.44	.26	—	.47	3.46
	(27,27,—,36)	110	115		0.47				(R = .65)	0.40
						Urban Planning				
M	School 145	494	519	—	3.15	.27	.29	—	.54	3.35
	(27,27,—,27)	111	142		0.40				(R = .58)	0.32
						Social Work				
M	School 293	540	489	—	—	.43	.14	—	—	3.42
	(146,146,—,—)	95	103							0.36

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

[a] Indicates degree orientation of students in the sample: M&D = both prospective master's and doctoral students, M = master's only.

Table 9

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

PSYCHOLOGY

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
D[a]	School 221 (45,45,33,42)	640 104	658 104	623 93	3.63 0.26	.42	.45	.38	.22 (R = .51)	3.14 0.55
M&D	School 145 (89,89,—,89)	663 96	655 118	—	3.62 0.34	.35	.35	—	.17 (R = .38)	3.62 0.34
M	School 122 (34,34,—,34,)	557 83	568 84	--	3.03 0.43	.20	.15	—	.32 (R = .44)	3.10 0.71
M	School 087 (25,25,21,—)	652 107	619 94	640 81	—	-.08	-.07	.17	—	3.53 0.39
M	School 115 (47,47,46,—)	511 86	548 95	—	3.40 0.40	.44	.21	—	.52 (R = .60)	3.72 0.25
M	School 068 (27,27,—,—)	567 105	544 103	—	—	.18 (Criterion = GPA Gen + Critical)	.21	--	—	3.65 0.48
?	School 095 (22,22,22,22)	609 68	614 91	609 58	3.20 0.41	.07	.19	.17	.18	3.62 0.32
D	School 297 (20,20,--,—)	621 60	627 100	—	—	-.36	.19	—	—	3.70 0.26
M&D	School 097 (76,76,76,—)	621 65	648 95	639 72	—	.14	.11	.25	—	3.82 0.17
D	School 069 (32,32,30,22)	666 70	655 64	605 80	3.43 0.26	.22	.48	.71	-.01	3.54 0.30
D	School 132 (53,53,47,51)	622 101	617 111	605 90	3.42 0.40	.52	.54	.48	.11 (R = .64)	3.58 0.33
D	School 203 (51,51,51,—)	598 72	581 84	604 59	—	.06 (Average of GPA and Dept. Rating)	.13	.42		3.58 0.60
M/D	School 123 (27,27,13,26)	548 81	564 77	576 79	3.28 0.41	-.16	-.31	.22	.29	3.89 0.16

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

[a] Indicates degree orientation of students in the sample: M&D = both prospective master's and doctoral students, D = doctoral candidates, M = master's candidates, ? = not indicated by department, M/D = master's plus some post-master's doctoral students.

Table 10

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

EDUCATION AND LIBRARY SCIENCE

	School (coded)	Predictor mean & s.d.				Validity coefficient				Grad GPA
	(N / coefficient)	GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	mean & s.d.
M[a]	School 007 (41,41,—,41)	446 69	461 86	—	2.89 0.37	Vocational & Adult Ed .16	.01	—	.24 (R = .33)*	3.75 0.20
M	School 069 (44,44,27,25)	465 89	460 85	473 63	3.06 0.40	Education .42	.19	.53	.46 (R = .60)	3.70 0.21
M	School 122 (70,70,—,—)		435** 96	—	—	Special Ed	.20**	—	—	3.52 0.61
M&D	School 145 (50,50,—,49)	526 116	521 131	—	3.22 0.47	Education .33	.22	—	.36 (R = .39)	3.60 0.40
M&D	School 231 (82,82,32,190)	527 122	519 123	525 114	3.11 0.39	Education .16	.26	.54	.19 (R = .57)	3.68 0.25
M	School 297 (36,36,—,—)	430 67	473 90	—	—	Physical Ed -.22	-.29	—	—	3.58 0.30
M	School 325 (27,27,—,27)	424 78	471 93	—	2.78 0.41	Administration and Supervision .13	.11	—	.18	3.60 0.36
M&D	School 332 (12,12,—,—)	467 96	514 66	—	—	Vocational-Technical Ed .25	.23	—	—	3.73 0.31
M	School 028 (27,28,—,28)	528 97	456 125	—	3.12 0.39	Library Science -.14	.30	—	.33	3.76 0.27
M&D	School 145 (51,51,—,52)	594 99	523 123	—	3.42 0.42	Library Science .46	.59	—	.46 (R = .77)	3.32 0.32
M	School 221 (38,38,—,38)	610 109	535 121	—	3.28 0.67	Library Science .47	.59	—	.15 (R = .63)	3.07 0.32

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Cross-validated composite of predictors.
**GRE Total (V+Q/2) only was reported.

[a] Indicates degree orientation of students in the sample: M = master's, M&D = both prospective master's and doctoral students.

Table 11

Summary of Validity Coefficients for GRE and UGPA Predictors versus First-Year Graduate GPA

LANGUAGES, MUSIC, PHILOSOPHY

	School (coded) (N / coefficient)	Predictor mean & s.d.				Validity coefficient				Grad GPA mean & s.d.
		GRE-V	GRE-Q	GRE-Adv	UGPA	GRE-V	GRE-Q	GRE-Adv	UPGA	
						French				
M&D[a]	School 221 (19,19,19,14)	569 125	502 120	550 63	3.59 0.28	.20	.43	.30	.30	3.17 0.27
						Spanish				
M&D	School 221 (14,14,14,14,)	467 151	446 110	562 102	3.36 0.49	.53	.70	.57	.26	2.79 0.56
						German				
M&D	School 204 (10,10,—,—)	558 103	471 79	—	—	.37	-.37	—	—	3.68 0.29
						Spanish				
M&D	School 204 (8, 8,—,—)	399	339	—	—	(+)*	(-)*	—	—	3.67 0.48
						Foreign Lang & Lit				
M&D	School 293 (37,37,—,—)	571 111	539 126	—	—	.25	.10	—	—	3.59 0.33
						Music				
M	School 069 (41,41,38,33)	505 101	516 113	507 74	3.23 0.37	.51	.38	.18	.23 (R = .53)	3.70 0.26
						Music				
M	School 123 (37,37,13,—)	477 121	482 75	512 106	—	.11	.09	.29	—	3.65 0.32
						Music				
M&D	School 204 (66,66,—,—)	541 132	515 147	—	—	14	-.04	—	—	3.83 0.34
						Philosophy				
D	School 035 (.9, 9, 9,—)	668 32	700 99	771 81	—	(-)*	(-)*	(-)*	—	3.81 0.22
						Philosophy				
M&D	School 101 (16,16,—,16)	675 60	644 72	—	3.58	.57	-.46	—	.24	3.25 0.33
						Philosophy				
M&D	School 145 (24,25,—,24)	693 59	637 86	—	3.66 0.33	.29	.01	—	.77	3.44 0.57
						Philosophy				
M&D	School 204 (17,17, 9,—)	639 95	549 103	691 99	—	.42	.22	(+)*	—	3.71 0.27
						Philosophy				
D+	School 231 (16,16,—,—)	654 83	628 136	—	—	.26	.50	—	—	2.93 0.78
						Music				
M+	School 231 (34.34.34.--)	521 93	485 97	517 69	--	.29	.02	.08	--	3.59 0.28

Note: Multiple correlation coefficients (R values in parentheses in UGPA column) are not reported in all instances due primarily to sample size and/or missing data considerations.

*Coefficient based on less than 10 cases; sign indicates direction of relationship.

[a] Indicates degree orientation of students in sample: M&D = both prospective master's and doctoral students, M = master's students, D = doctoral students, D+ = primarily prospective doctoral students with several master's students, M+ = primarily master's students with several prospective doctoral students.

Appendix D

Examination of Departmental Samples with Deviant Weights for Predictors

Appendix D

Examination of Departmental Samples With Deviant Weights for Predictors

The test results summarized in Table 16 indicate that the data conform generally to the common-weights hypothesis. However, what about the deviant departments? What characteristics in the data may be associated with the observed outcomes? To shed light on these questions, a detailed examination was made of the data for all departments in which one or more predictors was identified as having a slope differing significantly from the pooled estimate.

Table D-1 shows zero-order validity coefficients for V, GRE-Q, and UGPA, as available, for 12 departmental samples in which some departure from the common slopes hypothesis was indicated. Slopes for the variables that are specially marked were different from pooled estimates in the V, Q, and/or the V, Q, UGPA analyses. It may be noted that of the 15 marked coefficients, seven are positive and eight are negative. In almost every instance, examination of the original data for the samples involved revealed certain conditions that help to account for either the unusually high positive coefficients (and the correspondingly large positive regression weights) or the theoretically anomalous negative coefficients (and the corresponding negative regression weights).

Detail regarding each sample is provided following a general summary of the basic patterns of findings regarding the deviations:

- Positive deviations. In four samples characterized by unusually high positive weights for GRE-V and/or GRE-Q, the observed result is associated with one or more atypical data sets (in the comparatively small samples) for individuals who are in certain identifiable ways "atypical"--i.e., members of minority groups with a very low test score and also very low graduate school grades, and/or foreign students for whom English may not be the native language.

  In the sample of 11 students for Biology (B)[*], for example, one student (foreign by inference from name) presented a GRE-V score of 320 (2.51 standard deviation below the mean) and earned a GPA of 3.07 (1.84 S.D.'s below the mean). Without this data set, the sample coefficient would have been .42 rather than .80. The "inflated" GRE-V coefficient for Chemistry (F) is accounted for by one data set (foreign), while the unusually high GRE coefficients for Zoology (J) and Psychology (I) are heavily influenced by data sets for two and eight minority students, respectively, with atypically low GRE scores and graduate GPA.

---

[*] Departments are identified by letter in the detailed descriptions which follow. Letters are as indicated in Table D-1.

Table D-1

Validity Coefficients for Predictors in Departmental Samples with Deviant Weights on Designated Predictors

Department (School)		N	Validity coefficient for predictor		
			GRE-V	GRE-Q	UGPA
Chemistry	(A)	52	-.21*	.17	.37
Chemistry	(F)	13	.40*	.15	.03
Biology	(B)	11	.80**	-.03	
Botany	(C)	15	.32	-.43**	
Biology	(D)	26	.05	-.07#	.16
Zoology	(J)	19	.60	.72##	-.13
Psychology	(H)	20	-.36**	.19	
Psychology	(E)	27	-.16***	-.31***	.29
Psychology	(I)	51	.52*	.54##	.11
Psychology	(G)	46	.44	.21	.52**
History	(A)	46	.29	.22	.53**
History	(C)	25	.08	-.39***	-.01**
English	(None)				

Note: If the coefficient for a predictor is not specially marked, the corresponding weight was not identified as differing significantly from the pooled estimate.

*Corresponding weight deviant in analysis with Q constant but not with Q, U constant

**Corresponding weight deviant

***Corresponding weights deviant in analyses with V, and with V, GPA

#Corresponding weight deviant with V, UGPA constant, but not with V only

##Corresponding weight deviant in analysis with V, but not with V, UGPA constant

No specific sample characteristics could be identified to help "explain" the atypically high positive coefficients for UGPA in two samples.

- **Negative deviations**. In samples with anomalous negative coefficients, outcomes were associated with one or more of the following conditions: (1) one or two extremely atypical data sets or outliers [Chemistry (A), Botany (C), Psychology (H), History (C)] which heavily influenced results; (2) extremely skewed grade distribution [see especially Psychology (E)]; (3) minority and/or foreign student [see Chemistry (A), History (C)]; or (4) confounding interactions between level of GRE-scores, level of GPA, and year of entry [Biology (D)].

Detailed Description of Departmental Data

It is helpful first to consider the samples in which an atypically large positive weighting was present and then those in which large negative weighting was present for one or more predictors.

Positive Deviations

Chemistry (F): Sample of N = 11 includes two students, who by name identification appear to be foreign, with extremely low GRE-V scores and quite low graduate GPA (V = 290, GPA = 2.04 on 4.00 scale; V = 310, GPA = 3.18). GRE-Q score is in average range.

Biology (B): Sample of 11 students includes one foreign student (by name ID) with very low GRE-V (320) and very low GPA (3.07).

Zoology (J): Sample N = 19. Two minority students in sample, with data as follows:

V	Q	UGPA	Graduate GPA	
370	420	3.61	2.75	(Student 1)
430	410	3.04	2.57	
564	621	3.28	3.65	Dept. mean
96	94	0.35	0.40	Dept. sigma

High positive weights for V and Q are due primarily to positive covariation contributed by these two cases. Negative weight for UGPA (not identified as significantly deviant) is accounted for primarily Student 1.

Positive Deviations (cont.)

Psychology (I): As for Zoology in School J, the unusually high positive coefficients (weights) for V and Q were due to the fact that minority students with low GRE scores also tended to earn low graduate grades (i.e., correlated means).

Psychology (G):
Psychology (A): No specific patterns such as those outlined above can be identified to account for the high positive contribution of the UGPA in these two samples.

Negative Deviations

Chemistry (A): Correlation of GRE-V with Graduate GPA is -.21 in this sample (N = 52). One of three minority students (Oriental-American) had a GRE-V score of 310 (-4.21 standard deviations) and a Graduate GPA of "A" (+2.84 standard deviations). Elimination of this data set yields a sample in which the coefficient for GRE-V is .02. This individual's Q score was 760 (+1.00 S.D., approximately).

Botany (C): GRE-Q was negatively weighted in this sample (N = 15, Mean Q = 591, S.D. = 89). Two students with highest Q score, both 720, had the lowest and second lowest Graduate GPA (3.10 and 3.20) in group with Mean GPA = 3.70, S.D. = 0.30. No ethnic or language data were coded for this sample.

Psychology (H): Negative weighting was observed for GRE-V in this sample of 20 cases. Mean V = 621, S.D. = 60; mean Graduate GPA = 3.70, S.D. = 0.24. Individual with lowest Graduate GPA (3.19, or -2.13 S.D.) was one of two individuals with highest V score (740, or +1.98 S.D.). The other student with V = 740 earned GPA somewhat below average. Both these students had below average Q scores. No ethnic or language data were coded for this sample.

Psychology (E): Coefficients for V and Q were negative in this sample of 27 cases:

	Graduate GPA	
GRE-Verbal	Less than 4.00	4.00
600+	5	5
Less than 600	5	12
Total	10	17

Negative Deviations (cont.)

The Graduate GPA distribution for this department was heavily skewed negatively:

		Freq.	
(A average)	4.00	17	
	3.80	6	
	3.70	1	Mean = 3.89
	3.60	1	S.D. = 0.16
	3.50	2	

It seems probable that differences among students are not reliably measured.

History (C): In this department of 21 cases, Q and UGPA were identified as deviant (negative zero-order correlation with Graduate GPA). Student with lowest Q (310, -2.22 S.D.) earned 4.00 Graduate GPA (+1.25 S.D.); student with highest Q (650, + 1.69 S.D.) had lowest Graduate GPA (2.80, - 1.75 S.D.). Sample heterogeneity involved--several foreign students.

Biology (D): Negative weight for Q in this departmental sample (N = 26):

GRE-Q	GPA distribution		
	Below 3.25	3.25-3.74	3.75+
700+	4*	2	3
600-699	2	7	2
Below 600	1	4	1

*Cases in this cell account for negative correlation.

In this department, students entering in 1975 had lower GRE scores than those entering in 1974 but higher mean GPA. Relationship among year of entry, GRE variables and Graduate GPA was as shown below:

	Year 1975 vs. 1974	Graduate GPA
GRE-V	-.37 ('75 lower)	.05
GRE-Q	-.24 ('75 lower)	-.07
Grad GPA	.18 ('75 higher)	--

Some Implications of the Findings

These findings point up the impact of one or two aberrational data sets, or outliers, on the magnitude and the signs of validity coefficients in small samples. The negative coefficients, of course, are anomalous--i.e., coefficients reflecting the relationship between GRE and UGPA predictors, on the one hand, and first-year Graduate GPA, on the other, should be positive, a priori. Given the potential for anomalous "outlier" impact in small samples, the overwhelmingly positive distribution of coefficients obtained in the Cooperative Studies for GRE and UGPA in departmental samples with very small Ns, on the average, indicates a remarkable degree of underlying "regularity" in such data. Careful attention to sample definition clearly is important.

GRE BOARD RESEARCH
REPORTS OF A TECHNICAL NATURE

Boldt, R. R. Comparison of a Bayesian and a Least Squares Method of Educational Prediction. GRE Board Professional Report GREB No. 70-3P, June 1975.

Campbell, J. T. and Belcher, L. H. Word Associations of Students at Predominantly White and Predominantly Black Colleges. GRE Board Professional Report GREB No. 71-6P, December 1975.

Carlson, A. B.; Reilly, R. R.; Mahoney, M. H.; and Casserly, P. L. The Development and Pilot Testing of Criterion Rating Scales. GRE Board Professional Report GREB No. 73-1P, October 1976.

Carlson, A. B.; Evans, F.R.; and Kuykendall, N. M. The Feasibility of Common Criterion Validity Studies of the GRE. GRE Board Professional Report GREB No. 71-1P, July 1974.

Donlon, T. F.; Reilly, R. R.; and McKee, J. D. Development of a Test of Global vs. Articulated Thinking: The Figure Location Test. GRE Board Professional Report GREB No. 74-9P, June 1978.

Echternacht, G. Alternate Methods of Equating GRE Advanced Tests. GRE Board Professional Report GREB No. 69-2P, June 1974.

Echternacht, G. A Comparison of Various Item Option Weighting Schemes/A Note on the Variances of Empirically Derived Option Scoring Weights. GRE Board Professional Report GREB No. 71-17P, February 1975.

Echternacht, G. A Quick Method for Determining Test Bias. GRE Board Professional Report GREB No. 70-8P, July 1974.

Evans, F. R. The GRE-Q Coaching/Instruction Study. GRE Board Professional Report GREB No. 71-5aP, September 1977.

Fredericksen, N. and Ward, W. C. Development of Measures for the Study of Creativity. GRE Board Professional Report GREB No. 72-2P, June 1975.

McPeek, M.; Altman, R. A.; Wallmark, M.; and Wingersky, B. C. An Investigation of the Feasibility of Obtaining Additional Subscores on the GRE Advanced Psychology Test. GRE Board Professional Report GREB No. 74-4P, April 1976.

Powers, D. E.; Swinton, S.; Thayer, D.; and Yates, A. A Factor Analytic Investigation of Seven Experimental Analytical Item Types. GRE Board Professional Report GREB No. 77-1P, June 1978.

Powers, D. E.; Swinton, S. S.; and Carlson, A. B. A Factor Analytic Study of the GRE Aptitude Test. GRE Board Professional Report GREB No. 75-11P, September 1977.

Reilly, R. R. and Jackson, R. Effects of Empirical Option Weighting on Reliability and Validity of the GRE. GRE Board Professional Report GREB No. 71-9P, July 1974.

Reilly, R. R. Factors in Graduate Student Performance. GRE Board Professional Report GREB No. 71-2P, July 1974.

Rock, D. A. The Identification of Population Moderators and Their Effect on the Prediction of Doctorate Attainment. GRE Board Professional Report GREB No. 69-6bP, February 1975.

Rock, D. A. The "Test Chooser": A Different Approach to a Prediction Weighting Scheme. GRE Board Professional Report GREB No. 70-2P, November 1974.

Sharon, A. T. Test of English as a Foreign Language as a Moderator of Graduate Record Examinations Scores in the Prediction of Foreign Students' Grades in Graduate School. GRE Board Professional Report GREB No. 70-1P, June 1974.

Ward, W. C.; Frederiksen, N.; and Carlson, S. B. Construct Validity of Free-Response and Machine-Scorable Versions of a Test of Scientific Thinking. GRE Board Professional Report GREB No. 74-8P, November 1978.

Ward, W. C. and Frederiksen, N. A Study of the Predictive Validity of the Tests of Scientific Thinking. GRE Board Professional Report GREB No. 74-6P, October 1977.